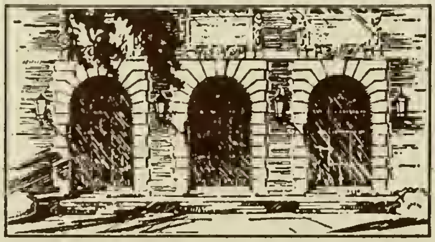


DIGITAL COMPUTER LABORATORY
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

REPORT NO. 136

STATIONARY DISTRIBUTION OF PARTIAL REMAINDERS IN S-R-T DIGITAL DIVISION

by

Richard Robert Shively

May 15, 1963

(This work is being submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Electrical Engineering, May 1963).

DIGITAL COMPUTER LABORATORY
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

REPORT NO. 136

STATIONARY DISTRIBUTION OF PARTIAL REMAINDERS IN S-R-T DIGITAL DIVISION

by

Richard Robert Shively

May 15, 1963

(This work is being submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Electrical Engineering, May 1963).

ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor J. E. Robertson for certain important observations, and for his encouragement and guidance in pursuing this work.

The author is also indebted to Professor G. A. Metze for his invaluable counsel at several points during this study.

The friendly help of Mrs. Phyllis Olson who typed the thesis in its final form is sincerely appreciated.

TABLE OF CONTENTS

STATIONARY DISTRIBUTION OF PARTIAL REMAINDERS IN S-R-T DIGITAL DIVISION

Richard Robert Shively

Department of Electrical Engineering

University of Illinois, 1963

The execution time required for digital division in electronic computers can be reduced if a) more than a minimum number of divisor multiples are available, and b) provisions exist to either by-pass or shorten steps which generate zero quotient digits. A division algorithm which uses this redundancy to increase the likelihood of zero digits (i.e., when the quotient is expressed in the recoded form implied by the divisor selections) then is desirable.

To provide a basis for comparing various division algorithms in this respect, C. V. Freiman observed, using a Markov chain model for the division process, that for any given divisor, D, and a randomly distributed set of dividends, the distribution of partial remainders at successive steps approaches a steady-state or stationary distribution, which is independent of the distribution of dividends. In attempting to determine this stationary distribution as a function of D for an algorithm known as S-R-T division, Freiman found that the pointwise solutions provided by a Markov chain are inadequate to analyze the function for certain intervals of D.

The purposes in further pursuing this work started by Freiman are:

a) to obtain a more complete understanding of S-R-T division,

b) to derive methods which might be found applicable in analyzing other division algorithms, and

c) to provide information which could prove useful in formulating new division algorithms.

If partial remainders, R_i, are defined to be in one-to-one correspondence with quotient digits (rather than considering only those associated with nonzero quotient digits), the stationary probability density, F(x), which

describes R_i, can be shown to be symmetric with respect to the point D, F(D), for all D in the allowed range, $\frac{1}{2} \leq D < 1$. This symmetry can be used to derive a straightforward method for evaluating the stationary density for intervals in the divisor range, rather than for individual points. The method essentially consists of locating points at which F(x) is discontinuous, as an ordered process, starting with points of discontinuity in the relation between R_i and R_{i+1}, and terminating when the first such discontinuity point falls in the interval $(\frac{1}{2}, 2D - \frac{1}{2})$. Then, by observing that the jump sizes at successive points in this ordered set of argument values form a geometric progression, and that the sum of these jumps is $\frac{1}{D}$, the solution follows directly. Solution intervals, i.e., neighborhoods in the range of D for which F(x) has similar form, may be grouped into infinite families, based on the sequence of operations used to locate discontinuity points, and when considered in the order of increasing complexity, these families form a nested pattern.

The stationary densities describing R_i can be used to determine the shift average, <S>, i.e., the average number of quotient digits formed per use of the adder. <S> is equivalent to the mean recurrence time of the event "$R_i \geq \frac{1}{2}$" in S-R-T division, and is therefore the reciprocal of the probability that $R_i \geq \frac{1}{2}$. Using properties of F(x), <S> can be shown to be $\frac{2D}{2D - 1}$ for $0.75 < D < 1$, reaching a maximum of 3, over the interval $0.6 < D < 0.75$. In the interval $0.5 < D < 0.6$, <S> requires an infinite number of subintervals to express completely, and exhibits a nonmonotonic behavior at certain points.

Analysis of other division algorithms can use the fact that a) the piecewise constant functions which describe the stationary density can be formed by locating discontinuity points as an ordered process, and b) the sizes of jumps at successive arguments in this process form a geometric progression.

1. BACKGROUND

1.1 Introduction

Numerous factors are effecting a willingness to invest in redundant equipment to improve digital computer operation times. Reliability and decreasing cost of components, modularization of circuits, and miniaturization are but a few of these. In arithmetic processes, some commonly considered applications of redundancy within the structure of a general purpose computer are: a) redundant number representation to reduce or eliminate adder carry propagation time, b) multiple shift paths to reduce the number of steps required in iterative processes such as multiplication and division, and c) generation of more than the minimum number of multiples of the operand, again to improve multiplication and division.

In digital division, redundancy in the generation of divisor multiples may be used to serve either or both of two general aims: a) providing overlap in the partial remainder ranges to which the respective divisor multiples are assigned, and b) minimizing the expectation value of the partial remainder magnitude. The first of these is particularly important when signed digits or stored carries are used, since distinguishing between nonoverlapping intervals by inspection of leading digits of unassimilated numbers would be impossible. Reduction in the partial remainder magnitude can, under certain circumstances, reduce the time required for the division process. (This will be discussed in Section 1.2.)

The effectiveness of a particular digital division algorithm (and the redundant equipment it may require) in reducing the average partial remainder, R_i,[1] can be measured if the probability distribution (or density) functions

[1] Hereafter, R_i will be used to denote the magnitude of the partial remainder occurring on the ith step of a division process.

which describe R_i are known. In Section 1.3 it will be shown, using the theory of finite Markov chains, that for a particular divisor, D, the distribution describing R_i converges with increasing i toward a unique steady-state distribution. Characterizing digital division as a discrete Markov process is due to C. V. Freiman (see Ref. 4). As Freiman discovered, however, the complicated behavior of the steady-state distribution of R_i, for certain ranges of the divisor, is not amenable to investigation using Markov chain techniques alone.

The purpose of the dissertation is to derive a method for more readily finding the steady-state distribution of R_i, assuming the S-R-T division algorithm is being used, and to apply this method in examining the function's behavior in those rangesof D where detailed analysis was previously impracticable. Other than satisfying a curiosity, the worth of this information has been demonstrated by Professor Gernot Metze[2] who, using properties indicated in previous efforts and proved herein, has proposed a modification to S-R-T division which yields, for all values of the divisor, D, the maximum shift average (i.e., maximum number of quotient digits per use of adder) obtainable under the constraint that the available divisor multiples are +D, 0, and -D.

1.2 S-R-T Division

Descriptions of conventional restoring and nonrestoring division can be found in various texts on digital computer arithmetic.[3]

By common usage, the phrase "S-R-T division" connotates what may be regarded as nonrestoring, binary division modified to utilize the facilities of

[2] See Ref. No. 7.

[3] e.g., Ref. No. 9.

a floating-point computer. Sweeney,[4] Robertson,[5] and Tocher[6] independently observed that if the absolute value of the fractional portion of the divisor, D, is restricted to the normalized range $.5 < D < 1$,[7] then any steps in which the partial remainder magnitude is found to have leading zeros can be by-passed, and zeros inserted as the corresponding quotient digits. (After being recoded from a representation using +1, 0, and -1 as digits to binary, the quotient digits just mentioned will be all zeros or all ones, depending on whether the next lower weighted nonzero quotient digit is positive or negative.) This is equivalent to saying we will always normalize the partial remainder between additions, and is profitable if paths to perform multiple shifts exist, or if the time for a given step is measurably dependent on whether or not that step requires use of the adder. Under these circumstances the average partial remainder magnitude is of interest, since the average number of leading zeros in R_i is equal to the average number of quotient digits formed per addition in the divison process.

To state the S-R-T division algorithm formally, the following definitions will be useful:[8]

a) P_i = the partial remainder after the ith step, where $i = 0, 1, \ldots, m-1$

b) $R_i = |P_i|$

c) D = divisor

[4] Ref. No. 2.

[5] Ref. No. 10

[6] Ref. No. 11

[7] Inclusion of end points is dependent on choice of number representation.

[8] In each case, the fraction portion of a floating-point number is implied.

d) q_i = the quotient digit generated by the ith step, with possible values of +1, 0, or -1

e) Q = quotient $\sum_{i=0}^{m-1} 2^{-i} q_i$

One way of stating the recursion relation is as follows:

1) If $R_i < \frac{1}{2}$, then $P_{i+1} = 2P_i$

$q_i = 0$

2) If $R_i \geq \frac{1}{2}$ and

a) the signs of P_i and D agree, then $P_{i+1} = 2(P_i - D)$

$q_i = 1$

b) the signs of P_i and D disagree, the $P_{i+1} = 2(P_i + D)$ and

$q_i = -1$

Since the ranges of $\frac{1}{2} < P_0 < 1$ and $\frac{1}{2} < D < 1$ would imply $\frac{1}{2} < Q < 2$, corrective initial and/or terminal steps may be required to obtain a quotient for which Q is a proper fraction. It can be proved by induction that the above process yields the correct quotient, since:

$$P_{i+1} = 2(P_i - q_i D) \tag{2.1}$$

and therefore after m steps are completed:

$$P_0 = 2^{-m} P_m + D \cdot \sum_{i=0}^{m-1} q_i \cdot 2^{-i} \tag{2.2}$$

or

$$P_0 = QD + 2^{-m} P_m \tag{2.3}$$

1.3 Division as a Markov Process

The partial remainder in a division process becomes a random variable if either or both of the dividend and divisor are random variables. A useful and somewhat surprising relation is that for any fixed divisor, the distribution of R_i is independent of the distribution of dividends, for i sufficientely large, provided variation in the dividend density is bounded.

1.3.1 Existence and Uniqueness of a Stationary Distribution

In order to demonstrate that R_i has a unique stationary distribution in the limit, Freiman[9] observed that the division process may be characterized as a finite, regular (or aperiodic and irreducible) Markov chain. The states of this Markov model are intervals in the range of R_i, each of which satisfies the following requirements:[10]

a) The probability density function describing R_i is uniform within the interval for all i. That is, if x_0 and x_1 are the end points of such an interval,

$$\Pr[x < R_i < x + dx] = K \cdot dx, \quad \text{for } x_0 < x < x_1 \qquad (3.1)$$

where K is a constant for any particular i, but may vary with i.

b) One such interval in the distribution of R_i will map into an integral number of intervals in the distribution of R_{i+1}, for all i. This means no new subintervals are required in order to insure that (a) is satisfied at each step.

[9] Reference 4.

[10] Proof that such states can be generated is deferred to Section 1.3.2.

Denote these states $\{S_j\}$, $j = 1, 2, \ldots, N$, and let $X(i)$, $i = 0, 1, 2, \ldots$, represent the stochastic process which ranges over this state space, where i corresponds to the step number, and $X(p) = q$ denotes the event "R_p is in state S_q." The Markovian character of $X(i)$ is a result of the two properties of all states listed above. If $X(i) = m$ is know to be true, the probability that $X(i + 1) = n$, for example, is not altered by information regarding which state $X(i - 1)$ assumed, because, due to the uniformity of the density function within an interval, this information would tell us nothing more about the probability that R_i is in the subset of S_m which maps into S_n. By induction, the random variable $X(i + 1)$ is also independent of all $X(p)$, $p \leq (i - 1)$, as well, or:

$$\begin{aligned} &\Pr[X(i + 1) = n|X(i) = m,\ X(i - 1) = k,\ \ldots,\ X(0) = a] \\ &\qquad = \Pr[X(i + 1) = n|X(i) = m] \end{aligned} \tag{3.2}$$

The rather trivial observation that the division algorithm is invariant from one step to the next enables further classifying the stochastic process as a stationary Markov chain, i.e., conditional probabilities $\Pr[X(i + 1) = n|X(i) = m]$ are transition probabilities which are independent of i. The procedure for finding the distribution of R_{i+1}, given that of R_i (or equivalently, finding the distribution of the discrete random variable $X(i + 1)$, given the distribution of $X(i)$) can be formally expressed by use of a transition matrix, P, with dimensions NXN for N states, and which has as its row m, column n entry the transition probability $\Pr[X(i + 1) = n|X(i) = m]$, i.e., rows are identified with sources, columns with destinations. Then if P is premultiplied by an N-entry row vector, $\pi(i)$, which describes the distribution of $X(i)$, i.e., element $\pi_{1m}(i) = \Pr[X(i) = m]$, the product is $\pi(i + 1)$. Since P is independent of i, this result can be generalized to:

$$\pi(i + k) = \pi(i) \cdot P^k; \quad k = 1, 2, \ldots \tag{3.3}$$

which expresses a k-step transition.

The desired result is that for i sufficiently large, $\pi(i)$ approaches a unique final distribution which is independent of $\pi(0)$. If λ denotes this limit distribution, then

$$\lambda = \lim_{k\to\infty}(\pi(0) \cdot P^k) \tag{3.4}$$

would be true for <u>any</u> 1 x N probability vector, $\pi(0)$. Equation (3.4) would obviously be true if each row of P^k approached λ as a limit, since the sum of the elements in $\pi(0)$ equals 1. We can prove that is precisely what the limiting behavior of P^k is, and can state as a

<u>Theorem</u>: If the entire interval $0 < R_i < 1$ can be covered with a finite number of intervals, each of which has the two requisite properties of a state listed at the start of 1.3.1, then there exists a unique vector λ such that equation (3.4) is true, for any $\pi(0)$; each row of P^k approaches λ as a limit.

<u>Proof</u>:

1. The conclusion holds if the chain is irreducible and aperiodic (e.g., see Feller, p. 356, or Kemeny, et. al., Chapt. 6).
2. Both properties are defined to apply if there exists a finite integer B, such that for all $k > B$ and all state pairs S_n, S_m, there is a nonzero probability of the system being in state S_m after k steps, having started in state S_n. (This applies to the case $n = m$ as well.) Equivalently, the properties apply if all elements of P^k are positive for $k > B$.

3. To show that B exists, we refer to equation (2.1), which, if attention is confined to absolute values, is represented in Figure 1.[11] The proof is completed with four observations:

   a) In the range of R_i, any interval of length $\leq \frac{1}{2}$ which does not include the points $R_i = .5$ or $R_i = D$, projects onto an interval twice its length in the range of R_{i+1}. The same comment applies to descendents, until one or both of .5 and D are covered.

   b) This doubling of the image with each step applies also to an interval which includes $R_i = .5$, provided $R_i = (D - .5)$ is not also included (see Figure 1). When both of these points are included, clearly D is included in the projected image on the range of R_{i+1}, to which observation (c) applies.

   c) An interval in the range of R_i containing D has a projected image expressible as $0 < R_{i+1} < \delta$, with $\delta > 0$; on successive iterations the image will be $(0, 2^n \cdot \delta)$, n = 1, 2, ... until the entire range of the partial remainder magnitude, (0,1), is covered.

   d) The first three observations established that, having started in any state (nonzero interval), the system can reach any other state, i.e., irreducibility. The fact that the last, <u>single</u> image of the sequence listed in (c) covers all of (0,1) proves the existence of the integer B, since this is equivalent to all elements

[11] Figure 1 derives from equation (2.1), together with $R_i = |P_i|$.

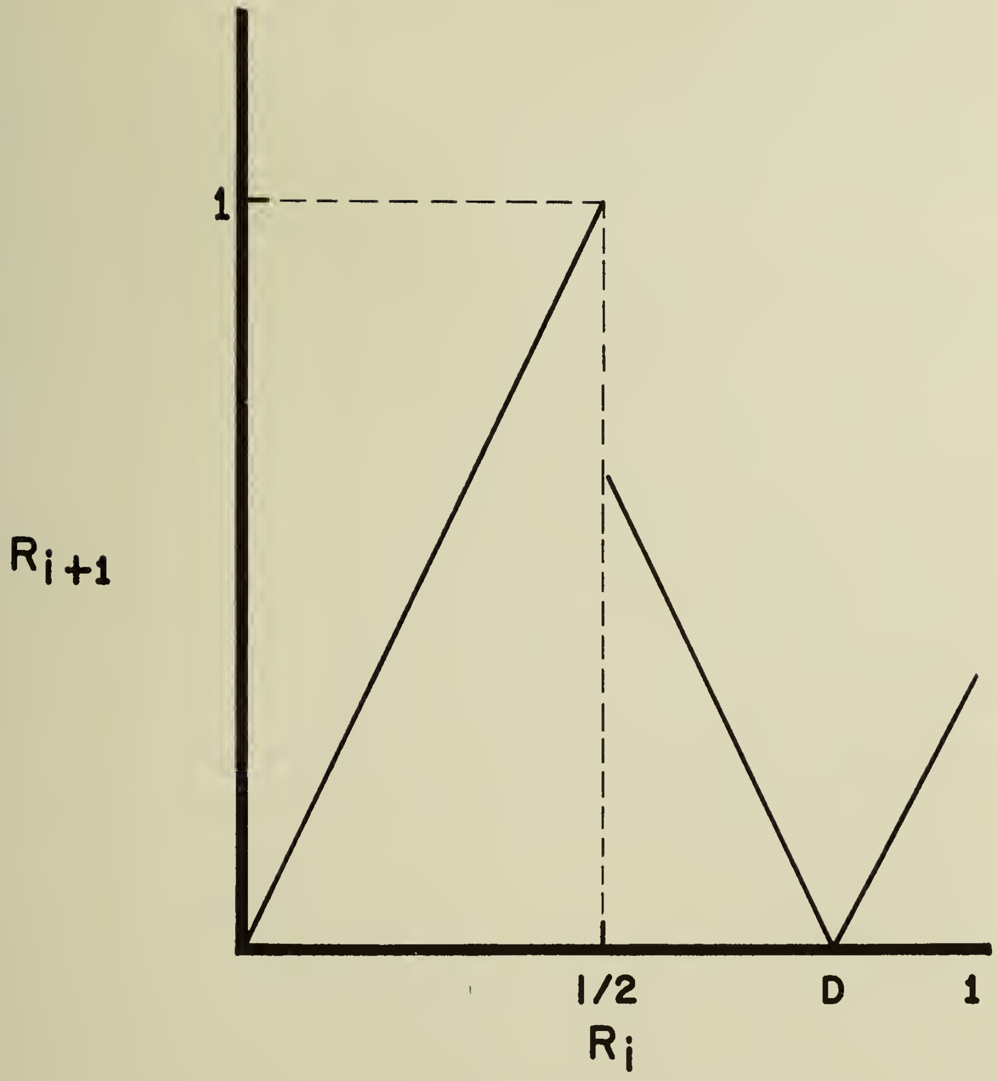


FIGURE 1: R_{i+1} vs. R_i

of P^k being positive for some k. Clearly all higher powers of P must also contain only positive elements, and hence the chain is aperiodic.

The last observation leads to an immediate

Corollary: All elements of λ are positive.

The fact that the stationary density function is nowhere zero in the unit interval will be used in Chapter 2.

1.3.2 Formulation of the Chain

First it will be shown that it is possible to generate a set of intervals covering (0,1) which meet the two requirements of states listed at the start of 1.3.1, for a large class of dividend distributions. With this established, the conditional hypothesis of the theorem in 1.3.1 can be removed, and for computational simplicity, the states may be selected on the assumption that the distribution of dividends is uniform, since the final result is independent of the actual distribution of dividends.

The states of the Markov chain for a given rational divisor, D, can be selected as follows:

a) While otherwise arbitrary, the density function describing dividend magnitudes is assumed to be sufficiently well behaved that it can be approximated in measure to any specified accuracy by a piecewise uniform density, with all discontinuities in the approximation occurring at rational numbers. (In particular, the density is everywhere finite; there is no single point in the range R_0 to which a non-zero probability is assigned.) Denoting the point set consisting of the divisor D and the arguments at which the discontinuities occur as $\left\{\frac{a_k}{b_k}\right\}$ (each point represented in reduced fractional form), we find the least common denominator. Denote this number as L.

b) The intervals, each L^{-1} in length, defined by the integral multiples of L^{-1}, are then the states. Due to the definition of L in (a), the density function for dividends is uniform within any such interval. Discontinuity arguments for the density functions of R_i, $i = 1, 2, \ldots,$ all are either images of these initial discontinuities, or of the points $R_0 = .5$ or $R_0 = 1$. (See Figure 1; the latter two are the points in the range of R_i at which the R_{i+1} versus R_i relation is discontinuous.) Since each such image is expressible in terms of integral multiples, and sums and differences, of elements in the set $\left\{\frac{a_k}{b_k}\right\}$, it is clear that density function uniformity within the specified intervals is preserved, and that the set of interval end points maps only onto itself.

Other than showing that the hypothesis in the theorem in 1.3.1 is satisfied, the above formulation reveals another interesting property of the stationary density function, viz., all discontinuities are images of the points .5 or 1, since the uniqueness property implies none of the discontinuities in the R_0 density and hence none of their images can influence the limiting density.

1.3.3 Examples

At this point two examples may be found helpful. We know that successive distributions of X(i) are related by the equation:

$$\pi(i + 1) = \pi(i) \cdot P \tag{3.5}$$

for all i, and that for i sufficiently large:

$$\pi(i) \doteq \pi(i + 1) \doteq \lambda \tag{3.6}$$

Hence, the solution for a given D requires finding P, then solving the equation set

$$\lambda = \lambda P \tag{3.7}$$

To determine state boundaries we generate the images of the points x = .5 and x = 1 by application of equation (2.1) or Figure 1, stopping when repetition begins. To determine the n,m element of the transition matrix, P, we answer the question: Given an ensemble of partial remainder uniformly distributed in state S_n, what fraction of them will be in S_m after one step?

Example 1: $D = \frac{7}{9}$

a) Images of the point $R_i = \frac{1}{2}$ are:

(i) $2(D - \frac{1}{2}) = \frac{5}{9}$

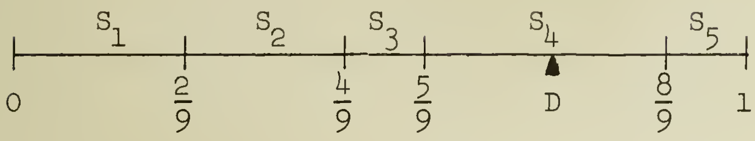
(ii) $2[D - (2D - 1)] = 2 - 2D = \frac{4}{9}$

(iii) $2[2 - 2D] = 4 - 4D = \frac{8}{9}$

(iv) $2[4 - 4D - D] = 8 - 10D = \frac{2}{9}$

Since the next application of equation (2.1) yields $\frac{4}{9}$, which is also the first image point of 1, all potential locations of discontinuities have been determined.

The states are therefore:



b) The transition matrix then is:

$$P = \begin{bmatrix} .5 & .5 & 0 & 0 & 0 \\ 0 & 0 & .25 & .75 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ .6\underline{6}\ldots & .3\underline{3}\ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

c) Solving the equation set $\lambda = \lambda P$, we obtain:

$$[\lambda_{11} \lambda_{12} \lambda_{13} \lambda_{14} \lambda_{15}] = [\tfrac{2}{7}\ \tfrac{2}{7}\ \tfrac{1}{7}\ \tfrac{3}{14}\ \tfrac{1}{14}]$$

Example 2: $D = \frac{3}{5}$

a) Images of .5 are:

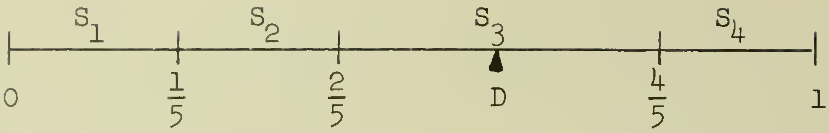
(i) $2(D - \frac{1}{2}) = \frac{1}{5}$

(ii) $2(\frac{1}{5}) = \frac{2}{5}$

(iii) $4(\frac{2}{5}) = \frac{4}{5}$

(iv) $2(\frac{4}{5} - D) = \frac{2}{5}$

The first image of 1 is also $\frac{4}{5}$. The states are therefore:



b)
$$P = \begin{bmatrix} .5 & .5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ .5 & .25 & 0 & .25 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

c) $\therefore [\lambda_{11}\ \lambda_{12}\ \lambda_{13}\ \lambda_{14}] = [\frac{1}{3}\ \frac{1}{4}\ \frac{1}{3}\ \frac{1}{12}]$

1.3.4 Limitations of the Markov Chain Approach

The Markov model has been very useful both conceptually and as a vehicle for proving the existence and uniqueness of a stationary probability distribution for partial remainders. However, when attempting to use the Markov chain to solve for this distribution, two significant obstacles are encountered:

a) The size of the linear equation set to be solved is governed by the value of the denominator of the divisor, D, rather than by the complexity of the density function being sought. This is a result of the definition of states which requires the set of end points to be closed under the recursion relation. Thus, in example 1 above, we find by normalizing the probability of each interval with respect to the length of the interval, the probability density function has only one discontinuity, although five equations were required (see Figure 2).

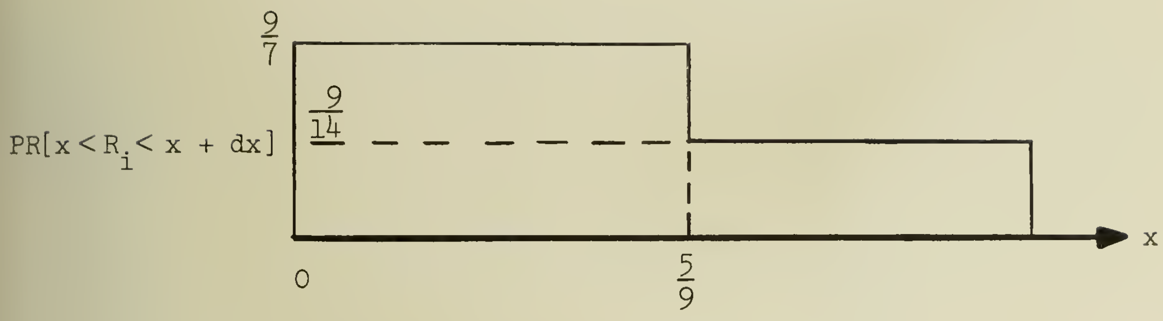


FIGURE 2: R_i STATIONARY DENSITY FOR $D = \frac{7}{9}$

b) Much more fundamental is the fact that a solution for a given value of D may reveal little about the behavior of the density function in that neighborhood of the divisor range. Thus it happens that in example 1 if we change D to any other value in the interval $(\frac{3}{4}, 1)$, the stationary density still has only one discontinuity, as will be shown in Chapter 2. However, if D is varied even an infinitesimal distance from $D = \frac{3}{5}$ in example 2, the number of discontinuities in the resulting density function may become arbitrarily large.

In Chapter 2, the equation which defines the steady-state distribution is examined. By deriving from it certain properties which must characterize the R_i stationary density functions for all values of D, a method to systematically generate solutions for intervals of D, without the use of simultaneous linear equations is developed.

2. DERIVATION OF PARTIAL REMAINDER DENSITIES

2.1 The Defining Equation

The distribution of partial remainders at the i+1 step of an S-R-T division process may be expressed in terms of the distribution at the preceding step as follows:

$$f_{i+1}(x) = \begin{cases} f_i(\frac{x}{2}) + f_i(D + \frac{x}{2}) - f_i^*(D - \frac{x}{2}), & 0 < x < 1 \\ 1, & x > 1 \\ 0, & x < 0 \end{cases} \tag{1.1}$$

where:

a) $f_i(y) \equiv \Pr[R_i < y]$, for a given divisor D;

b) $$f_i^*(y) \equiv \begin{cases} f_i(y), & y \geq \frac{1}{2} \\ f_i(\frac{1}{2}), & y < \frac{1}{2} \end{cases} \tag{1.2}$$

c) $\frac{1}{2} < D < 1$

This relation can be demonstrated graphically as in Figure 3, where the intervals of R_i which yield an R_{i+1} less than a specified number, x, are indicated. In the limit, if distributions of successive partial remainders are equal, let

$$f(x) \equiv \lim_{i\to\infty}[f_i(x)] \tag{1.3}$$

Then we have equation (1.1) with subscripts deleted:

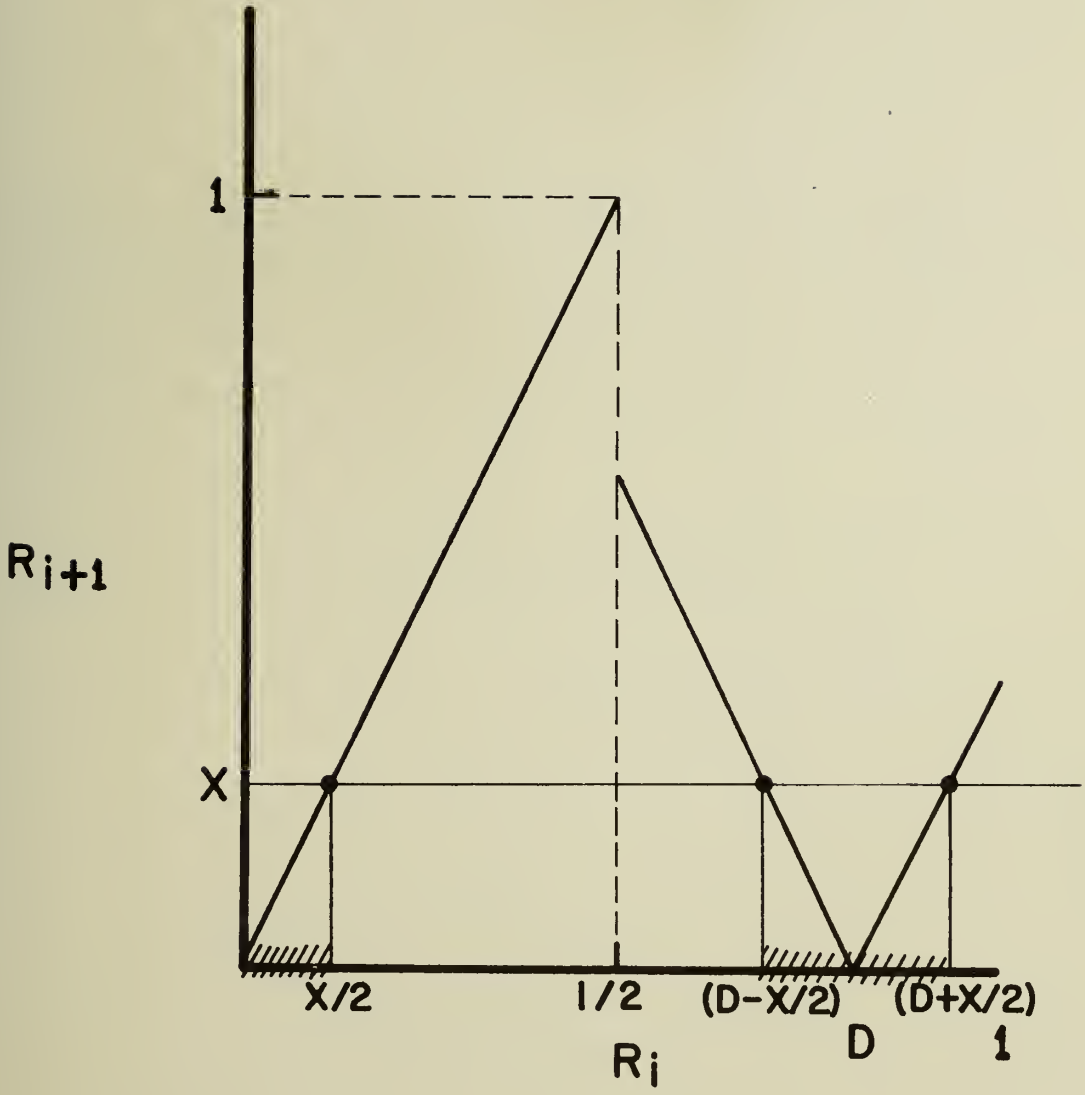


FIGURE 3: .f(x) = f(x/2) + f(D+x) -f(D-X/2)

$$f(x) = \begin{cases} f(\frac{x}{2}) + f(D + \frac{x}{2}) - f^*(D - \frac{x}{2}), & 0 < x < 1 \\ 1, & x > 0 \\ 0, & x < 0 \end{cases} \qquad (1.4)$$

to which (1.2) applies, again with subscripts deleted.

Because attention is restricted to piecewise uniform distributions, i.e., piecewise linear distribution functions, $f(x)$ is differentiable at all but a finite number of points. Therefore, to find the density function, we can take the derivative of both sides of equation (1.4) (to simplify notation, let $F(x) \equiv f'(x)$):

$$F(x) = \begin{cases} \frac{1}{2}[F(\frac{x}{2}) + F(D + \frac{x}{2}) + F^*(D - \frac{x}{2})], & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (1.5)$$

where

$$F^*(y) = \begin{cases} F(y), & y > \frac{1}{2} \\ 0, & y < \frac{1}{2} \end{cases}$$

Since $F(x)dx$ has the interpretation: $\Pr[x < R_i < x + dx]$, equation (1.5) is expressing the probability of a partial remainder being in any differential interval, say dx_0, as the sum of probabilities that its predecessor was in one of the (at most three) mutually exclusive differential intervals, each half as long as dx_0, which map into dx_0. Hence we have, in terms of a function whose domain is a continuum,[1] the relation that was expressed in terms of discrete state probabilities by equation 1:(3.7). The theory of Markov chains can be applied to demonstrate that a piecewise constant density function which satisfies (1.5) for a given D is unique. As in Chapter 1, generality will be slightly compromised by assuming D is rational. The first theorem below is instrumental in the proof of the second.

Theorem 2.1a): For D rational, and F(x) a piecewise constant density function satisfying (1.5), all discontinuities in F(x) occur at rational values of x.

Proof:

1. To construct a contradiction, assume there exists an irrational number, x_0, at which the conclusion is violated.
2. Since the number of discontinuities occurring at irrational arguments must be the same for both sides of (1.5), jumpsin two different terms on the right of that equation cannot cancel each other for x irrational. (As demonstrated later, such mutual cancellation can occur for x rational.) Hence, (1.5) implies that a second irrational discontinuity point exists at one of $2x_0$, $2D - 2x_0$, or $2x_0 - 2D$, depending on which of $(0, \frac{1}{2})$, $(\frac{1}{2}, D)$, or (D, 1) includes x_0.
3. By induction, the second such discontinuity implies the existence of a third, etc. The n'th term in the sequence is obviously expressible as:

$$x_n = \pm 2^n \cdot x_0 + C \qquad (1.6)$$

where C is some rational number (because D is rational).

4. Since a piecewise constant function is defined to have at most a finite number of discontinuities, the sequence described by (1.6) must cycle.

[1] In the mathematical model, representation of any partial remainder, rational or irrational, is assumed possible.

If x_0 is a point in this cycle, and recurs after k steps, we could infer from (1.6):

$$x_0 = \frac{C}{1 \pm 2^k}$$

which is impossible with x_0 irrational.

Theorem 2.1b): There is one, and only one, function which satisfies equation (1.5) for a given rational divisor, D, under the implied restrictions of a piecewise uniform probability distribution: (i) $\int_0^1 F(x)dx = 1$, (ii) F(x) is piecewise constant, and (iii) $F(x) \geq 0$ for all x.

Proof:

1. At least one solution exists, since the vector representing state probabilities of a Markov chain (i.e., the solution to equation 1:(3.2)), can alternatively be expressed as a density which must satisfy (1.5) at each point.
2. Conversely, a solution to (1.5) for a given divisor must be unique if the function can alternatively be expressed in terms of state probabilities of a finite[2] Markov chain, i.e., if there exists a partitioning of the remainder range such that the distribution is uniform within each interval, and the finite partition-point set is closed under the S-R-T recursion relation. Then, for two alleged solutions to (1.5) with D fixed, we could select a single partitioning (viz., the union of two partition point sets which apply to the solutions individually) for both functions, and thereby construct a contradiction based on the uniqueness of regular Markov chain state probabilities.

[2] While this uniqueness extends to regular Markov chains with a denumerably infinite number of states, the a priori knowledge that the number of discontinuities is finite, gained by restricting D to the set of rational numbers, is useful in the subsequent development (see Ref. 1, Chapt. 1 for properties of chains with an infinite number of states).

3. To select the partition point set hypothesized in 2, we observe that the number of discontinuities is finite, and all occur at rational numbers due to theorem 2.1a. Hence, the set $\left\{\frac{n}{L}\right\}$, n = 0, 1, ..., L, where L is the least common denominator of the divisor and discontinuity arguments, is always sufficient for the partition point set described in step 2 above. Q.E.D.

The problem of finding the stationary density of partial remainders is now reduced to solving (1.5) for F(x), as a function of D. Since no single number in the range of partial remainders has a nonzero probability, the definition of F(x) for x = 0, 1, or any of the finite number of discontinuity arguments will not be a primary concern. Moreover, D will be restricted to the open interval, $\frac{1}{2} < D < 1$, both for expediency in derivations, and because F(x) for $D = \frac{1}{2}$ is known; to wit: $F(x) = 1,\ 0 < x < 1$.

The definition of a single step in the S-R-T division process used here does not agree with that used by Freiman.[3] Instead of defining a step to be either one binary shift (i.e., if $q_i = 0$) or one addition and shift, Freiman defines a step to be one addition and partial remainder normalization. If g(x) denotes the steady-state distribution of the normalized partial remainder magnitudes, it is apparent from what computer normalization entails that:

$$g(x) = \begin{cases} \sum\limits_{n=0}^{\infty} [f(\frac{x}{2^n}) - f(\frac{1}{2^n})]; & \frac{1}{2} < x < 1 \\ 0; \quad x < \frac{1}{2} \end{cases} \qquad (1.7)$$

from which we obtain:

[3] Ref. 4.

$$g'(x) = \begin{cases} \sum\limits_{n=0}^{\infty} \frac{1}{2^n} F(\frac{x}{2^n}); & \frac{1}{2} < x < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (1.8)$$

Hence, information obtained using one definition can be readily compared with results for which the other was used. However, two advantages were found in using the definition adopted here. One is the very useful symmetry which characterizes all solutions to (1.5), as proved in Section 2.3. The other is a simplification in evaluating the expectation value for the number of shifts occurring between additions. The latter point will be discussed in Chapter 3.

2.2 End Point Relations

Two properties of all solutions can be deduced by direct substitution into (1.5):

$$F(0+) = 2F(D) \qquad (2.1)$$

$$F(\tfrac{1}{2}-) = 2F(1-) \qquad (2.2)$$

where, for example,

$$F(0+) \equiv \lim_{\epsilon \to 0}[F(0 + \epsilon)]$$

In (2.1), F(D) has the interpretation:

$$\frac{1}{2}\,[F(D+) + F(D-)]$$

if D is itself the argument of a discontinuity. Both (2.1) and (2.2) depend on the fact that the number of discontinuities is finite, since the limits might not exist if 0, $\frac{1}{2}$, D or 1 were accumlation points of discontinuities. (2.2) also relies on the fact that $D = \frac{1}{2}$ is not being considered.

2.3 Symmetry

Theorem: All solutions to equation (1.5) are symmetric with respect to the point defined by coordinates D, F(D), the symmetry extending over the interval $0 < x < 2D$; i.e.,

$$F(D - \theta) - F(D) = F(D) - F(D + \theta), \qquad 0 \leq \theta < D$$

for all D in the $(\frac{1}{2}, 1)$ interval.

Proof:

To summarize what follows, F is extended to a function with period 2D, and it is shown that the even portion (excluding an additive constant) of the convergent Fourier series representation of this extension must sum to zero for all x, which, due to Fourier series uniqueness, implies each cosine term is identically zero. The conclusion is an immediate sequel.

1. Define

$$H(x) = \begin{cases} F(x), & 0 \leq x < 2D \\ F(x_{MOD.2D}), & \text{otherwise} \end{cases} \qquad (3.1)$$

Due to the fact that F(x) is a piecewise constant function for all D and finite for all x, F(x) has bounded variation. Therefore, the Dirichlet-Jordan theorem[4] applies when formulating a Fourier series representation of H(x). If we define:

[4] e.g., see Zygmund; Theorem 8.1, Chapt. II.

$$b_n = \frac{1}{D}\int_0^{2D} \sin(\frac{n\pi x}{D}) \cdot F(x)dx \tag{3.2}$$

$$a_n = \frac{1}{D}\int_0^{2D} \cos(\frac{n\pi x}{D}) \cdot F(x)dx \tag{3.3}$$

such that the series corresponding to H(x) is:

$$S[H(x)] = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left\{b_n \sin(\frac{n\pi x}{D}) + a_n \cos(\frac{n\pi x}{D})\right\} \tag{3.4}$$

then the Dirichlet-Jordan theorem states that S[H(x)] converges uniformly to H(x) within any closed interval in which H(x) has no discontinuities, and, moreover, if x_0 is the argument of a discontinuity,

$$S[H(x_0)] = \frac{1}{2}\left\{H(x_0+) + H(x_0-)\right\} \tag{3.5}$$

Therefore the notational distinction between H(x) and its series representation is unnecessary if we define H(x) (and F(x)) as the average of left- and right-hand limits at each jump.

2. Since the period is 2D, and there is unit area under F(x), we know the constant of the series:

$$\frac{a_0}{2} = \frac{1}{2} D \tag{3.6}$$

To evaluate a_n, n > 0, the expression for F(x) in (1.5) may be substituted into (3.3).

$$a_n = \frac{1}{2D}\int_0^1 \cos(\frac{n\pi x}{D}) \cdot F(\frac{x}{2})dx + \frac{1}{2D}\int_0^{2D} \cos(\frac{n\pi x}{D})F(D + \frac{x}{2})dx$$

$$+ \frac{1}{2D}\int_0^{2D-1} \cos(\frac{n\pi x}{D})F*(D - \frac{x}{2})dx \tag{3.7}$$

The limits of integration in (3.7) are determined by the intervals through which x may vary in the respective terms on the right of (1.5). The upper limit on the second integral may be any value between 2 - 2D and 2D inclusive; all other limits are fixed. The flexibility of this one limit is due to $H(D + \frac{x}{2})$ being zero for that interval of x, together with the fact that it was unnecessary to impose any nonlinearity on the argument of the $F(D + \frac{x}{2})$ term in (1.5). (The equation defining F(x) could alternatively be written as:

$$F(x) = \frac{1}{2}[F_A(\frac{x}{2}) + F(D + \frac{x}{2}) + F_B(D - \frac{x}{2})], \quad \frac{1}{2} < D < 1 \tag{3.8}$$

where

$$F_A(y) = \begin{cases} F(y), & y < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$$

$$F_B(y) = \begin{cases} F(y), & y > \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$$

no other restrictions on the range of x are necessary.)

Substituting the series representation for F in the integrands of (3.7), we obtain the following expressions:

a) First term on the right of (3.7):[5]

$$\frac{1}{2D}\int_0^1 \frac{1}{2D}\cos(\frac{n\pi x}{D}) + \frac{1}{2}\sum_{\nu=1}^{\infty}\left\{a_\nu[\cos(n+\frac{\nu}{2})\frac{\pi x}{D} + \cos(n-\frac{\nu}{2})\frac{\pi x}{D}]\right.$$

$$\left. + b_\nu[\sin(n+\frac{\nu}{2})\frac{\pi x}{D} + \sin(n-\frac{\nu}{2})\frac{\pi x}{D}]\right\} dx \; ^6 \tag{3.9}$$

$$= \frac{1}{4Dn\pi}\sin(\frac{n\pi}{D}) + \frac{1}{4D}a_{2n} + \frac{1}{4\pi}\sum_{\nu=1}^{\infty}{}' \left\{a_\nu\left[\frac{\sin(n+\frac{\nu}{2})\frac{\pi}{D}}{n+\frac{\nu}{2}} + \frac{\sin(n-\frac{\nu}{2})\frac{\pi}{D}}{n-\frac{\nu}{2}}\right]\right.$$

$$\left. + b_\nu\left[\frac{1-\cos(n+\frac{\nu}{2})\frac{\pi}{D}}{n+\frac{\nu}{2}} + \frac{1-\cos(n-\frac{\nu}{2})\frac{\pi}{D}}{n-\frac{\nu}{2}}\right]\right\}$$

where a prime on a summation symbol means the index, ν, assumes all positive integer values excluding 2n for terms in which $(n - \frac{\nu}{2})^{-1}$ is a factor. The same notation is used below.

b) Second term on the right of (3.7) is:

$$\frac{1}{2D}\int_0^{2D} \frac{1}{2D}\cos(\frac{n\pi x}{D}) + \frac{1}{2}\sum_{\nu=1}^{\infty}(-1)^\nu\left\{a_\nu\left[\cos(n+\frac{\nu}{2})\frac{\pi x}{D} + \cos(n-\frac{\nu}{2})\frac{\pi x}{D}\right]\right.$$

$$\left. + b_\nu\left[\sin(n+\frac{\nu}{2})\frac{\pi x}{D} + \sin(n-\frac{\nu}{2})\frac{\pi x}{D}\right]\right\} dx \tag{3.10}$$

$$= \frac{1}{2}a_{2n} - \frac{1}{4\pi}\sum_{\nu=1}^{\infty}{}' b_\nu\left\{\frac{1-(-1)^\nu}{n+\frac{\nu}{2}} + \frac{1-(-1)^\nu}{n-\frac{\nu}{2}}\right\}$$

[5] n is constant, ν is the index of the series.

[6] Trigonometric identities used are: $\cos A \cos B = \frac{1}{2}[\cos(A+B) + \cos(A-B)]$

$\cos A \sin B = \frac{1}{2}[\sin(A+B) + \sin(A-B)]$

c) Third term on the right of (3.7):

$$\frac{1}{2D}\int_0^{2D-1} \cos(\frac{n\pi x}{D}) + \frac{1}{2}\sum_{\nu=1}^{\infty}(-1)^{\nu}\left\{a_{\nu}\left[\cos(n+\frac{\nu}{2})\frac{\pi x}{D} + \cos(n-\frac{\nu}{2})\frac{\pi x}{D}\right]\right.$$

$$\left. - b_{\nu}\left[\sin(n+\frac{\nu}{2})\frac{\pi x}{D} + \sin(n-\frac{\nu}{2})\frac{\pi x}{D}\right]\right\} dx \tag{3.11}$$

$$= -\frac{\sin(\frac{n\pi}{D})}{4Dn\pi} + \frac{a_{2n}(2D-1)}{4D} + \frac{1}{4\pi}\sum_{\nu=1}^{\infty}{}' \left\{-a_{\nu}\left[\frac{\sin(n+\frac{\nu}{2})\frac{\pi}{D}}{n+\frac{\nu}{2}} + \frac{\sin(n-\frac{\nu}{2})\frac{\pi}{D}}{n-\frac{\nu}{2}}\right]\right.$$

$$\left. + b_{\nu}\left[\frac{\cos(n+\frac{\nu}{2})\frac{\pi}{D} - (-1)^{\nu}}{n+\frac{\nu}{2}} + \frac{\cos(n-\frac{\nu}{2})\frac{\pi}{D} - (-1)^{\nu}}{n-\frac{\nu}{2}}\right]\right\}$$

Summing the expressions developed in (3.9), (3.10), and (3.11), we find that (3.7) reduces to:

$$a_n' = a_{2n} \tag{3.12}$$

3. As mentioned earlier, the variation per period of H(x) is known to be bounded. For variation V per period of length 2D, it can be shown that:[7]

$$|a_n| \leq \frac{V}{nD} \tag{3.13}$$

Now, by induction, (3.12) implies that a_n can be equated to any term of the form: a_k, $k = 2^j \cdot n$, where j is an arbitrarily large integer. Hence, (3.12) and (3.13) together imply that every $|a_n|$, n = 1, 2, ..., is less

[7] e.g., see Zygmund, Theorem 4.12, Chapt. 2. He derives the bound on a_n for a function with period 2π, but the modifications to his derivation are straightforward.

than any specified positive number. The question of summing an infinite number of arbitrarily small numbers is resolved by applying Cauchy's condition for convergence, i.e., given $\epsilon > 0$, there exists an integer N such that all partial sums of the series after the Nth are within ϵ of the limit sum. Because $\sum_1^\infty a_n \cos(\frac{n\pi x}{D}) \leq \sum_1^\infty |a_n|$, and $|a_n|$ can be shown to be less than ϵ/N by use of (3.12) and (3.13), it follows that:

$$|\sum_1^\infty a_n \cos(\frac{n\pi x}{D})| \leq \epsilon \tag{3.14}$$

Since ϵ can be arbitrarily small, the cosine portion of the series must be zero for all x.

For any integer n, $\sin(\frac{n\pi x}{D})$ is symmetric with respect to D, and therefor any sum of terms of this form is also symmetric. Q.E.D.

The significance of being able to select 2D as the upper limit on the second integral in (3.7) is now apparent. If any other limit in the allowed range of [2 - 2D, 2D] were used, the expression for the unknown, a_n, would be in terms of an infinite number of unknowns, the a_ν and b_ν, which, though correct, would have presented a more formidable problem than equation (3.12).

2.4 Corollaries to Symmetry

The symmetry of F(x) with respect to D will be used frequently in the remainder of this chapter, as well as the next. The properties listed in this section are some of the more immediate consequences of symmetry.

Theorem 2.4a:

$$\text{(i)} \quad F(D) = \frac{1}{2D} \tag{4.1}$$

$$\text{(ii)} \quad F(x) = \frac{1}{D}, \qquad 0 < x < 2D - 1 \tag{4.2}$$

$$\text{(iii)}\quad F[(2D - 1)-] - F[(2D - 1)+] = F(1-) \qquad (4.3)$$

Proof:

1. The first assertion follows from the fact that all terms except the constant are zero on the right of (3.4) at $x = D$. The constant is given by (3.6). (Again, the average of left- and right-hand limits is implied if D is itself a discontinuity point.)

2. Since, for $0 \leq \theta < D$, symmetry implies:

$$F(D - \theta) - F(D) = F(D) - F(D + \theta) \qquad (4.4)$$

and therefore, due to (4.1):

$$F(D - \theta) + F(D + \theta) = \frac{1}{D}, \quad 0 \leq \theta < D \qquad (4.5)$$

equation (4.2) follows if we note that $F(D + \theta) = 0$ if θ is in the interval: $1 - D < \theta < D$.[8]

3. Equation (4.3) is a particular, but useful, example of symmetry.

From properties of solutions established up to this point, we can deduce the following about $F(x)$ for all D in the interval $\frac{3}{4} < D < 1$:

a) $F(\frac{1}{2}) = \frac{1}{D}$, due to (4.2);

b) $F(1-) = \frac{1}{2D}$, due to (2.2) and (4.6a); (4.6)

c) $F(D) = \frac{1}{2D}$, due to (4.1)

d) $F[(2D - 1)+] = \frac{1}{2D}$, due to (4.3) and (4.6b).

[8] The upper limit on θ is merely to confine attention to the interval to which symmetry applies.

Because $F(x)$ is equal to $\frac{1}{2D}$ at each end and the midpoint of the interval $2D - 1 < x < 1$ (for $\frac{3}{4} < D < 1$), the question naturally arises as to whether or not solutions for all D in $(\frac{3}{4}, 1)$ are constant over this interval. This question is one of many answered by the

Theorem 2.4b: $F(x)$ is monotonic for $x > 0$, i.e., there is no $x_0 > 0$ such that

$$F(x_0+) - F(x_0-) > 0 \qquad (4.7)$$

Proof·

(A discontinuity described by (4.7) will be termed a positive jump; if the inequality is reversed, it is a negative jump.)

1. Considered individually, the respective terms on the right of equation (1.5) would tend to yield a positive jump in $F(x)$, for $x = x_0$, if there were:

   a) a positive jump at $x_0/2$

   b) a positive jump at $D + x_0/2$, for $x_0 < 2 - 2D$, or

   c) a negative jump at $D - x_0/2$, for $x_0 < 2D - 1$.

   Due to (4.2), the last of these is impossible. Therefore, any positive jump implies the existence of another.

2. Both 1a) and 1b) above cannot be true for the same value of x_0, since the number of positive jumps on the right in (1.5) would not then be the sum of the numbers of such jumps in the individual terms, which would lead to an inconsistency, since (1.5) is expressing $F(x)$ in terms of itself.

3. Assume $x_0 > 0$ is the argument of the largest positive jump (i.e., the difference in (4.7) is maximum). The contradiction is immediate, since (1.5) together with the above comments require the existence of a positive jump twice as large. Hence, there are none.

2.5 Generation of Solutions

Due to equations (4.2) and (4.6), together with the theorem just proved, we can confirm that the stationary partial remainder density functions for all divisors in the range $\frac{3}{4} < D < 1$ are:

$$F(x) = \begin{cases} \frac{1}{D}, & 0 < x < 2D - 1 \\ \frac{1}{2} D, & 2D - 1 < x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

For $D < \frac{3}{4}$, we know that $x = 2D - 1$ is again the argument of a discontinuity, nonzero in size, since this jump is equal to that at $x = 1$, and the latter is nonzero since $F(x)$ is nowhere zero inside the unit interval. In (1.5), this discontinuity appears in the range of the $F(\frac{x}{2})$ term, for $D < \frac{3}{4}$, and cannot be cancelled by an equal but opposite jump in one of the other terms in that equation because $F(x)$ is monotonic. Therefore, there will be a discontinuity at $x = 4D - 2$; if $2D - 1 < \frac{1}{4}$, then discontinuities exist at $4D - 2$ and $8D - 4$; etc.

Similar reasoning applies to discontinuities at points in the $(2D - \frac{1}{2}, 1)$ interval. Due to symmetry, a discontinuity in the range of either $F(D + \frac{x}{2})$ or $F^*(D - \frac{x}{2})$ is cancelled by a compensating change in the other of these terms in (1.5), provided both terms are nonzero. However, for $x > 2D - 1$, i.e., $D + \frac{x}{2} > 2D - \frac{1}{2}$, the F^* term is zero; in this case discontinuities in the range of the $F(D + \frac{x}{2})$ term cannot be cancelled, again because $F(x)$ is monotonic. The following summarizes these observations:

Theorem 2.5a: A discontinuity in $F(x)$, for $x = y_0$, $0 < y_0 < \frac{1}{2}$, implies the existence of a discontinuity for $x = 2y_0$. Similarly, a discontinuity for

$x = y_1$, $2D - \frac{1}{2} < y_1 < 1$, implies the existence of a discontinuity at $x = 2y_1 - 2D$.

Another useful conclusion which follows almost directly from the above observations is:

Theorem 2.5b: There is at most one pair of discontinuities in $F(x)$ for the interval $\frac{1}{2} < x < 2D - \frac{1}{2}$.

Proof:

This restriction is simply a consequence of the fact that the number of jumps on both sides of (1.5) must be equal for $x > 0$. A pair of jumps in the $\frac{1}{2} < x < 2D - \frac{1}{2}$ interval cancel each other on the right of (1.5), due to symmetry. The two discontinuities in the defining equation for $F(x)$ (see equation (3.8)) are each the source of one negative jump,[9] and thereby compensate for the mutual cancellation, or "loss," of two other jumps on the right side, but no more.

Since the intersection of the domains of the three terms on the right of equation (3.8) is null, while the union covers the unit interval (excluding the points $x = D$, $x = \frac{1}{2}$), each discontinuity in $F(x)$ appears in precisely one of these individual terms. Due to the two additional discontinuities introduced by the definitions of F_A and F_B in (3.8), the number of discontinuities in $F(x)$, $x > 0$ is, in general, two less than the sum of the numbers of discontinuities in these individual terms. This difference can be accounted for either by the mutual cancellation of jumps in the $(\frac{1}{2}, 2D - \frac{1}{2})$ interval, described in theorem 2.5b, or by two pairs of jumps in the individual terms

[9] viz.: jumps at $x - 2D = 1$ and $x = 1$ due to the zero-going of $F(D - \frac{x}{2})$ and $F(\frac{x}{2})$, respectively.

summing to a single pair in F(x). The stationary density function for $D = \frac{3}{5}$ is an example of the latter (see Figure 4). Since discontinuity arguments are all linear expressionsin D (e.g., 2D - 1, 4 - 6D, etc.), this effective coincidence of two pairs of discontinuity arguments in general occurs only for isolated values of D, which will be called the "exceptional divisors," denoted $\{D_E\}$. To be precise, the set $\{D_E\}$ is defined to include all divisors to which (5.2d) below applies.

We are now in a position to find the arguments of all discontinuities in a very straightforward way. A sequence $\{x_i\}$ is formed, using theorems 2.5a and 2.5b, starting with $x_1 = 2D - 1$:

a) If $x_i < \frac{1}{2}$, then $x_{i+1} = 2x_i$.

b) If $x_i > 2D - \frac{1}{2}$, then $x_{i+1} = 2D - 2x_i$.

(5.2)

c) If $\frac{1}{2} \leq x_i < 2D - \frac{1}{2}$, the sequence is complete.[10]

d) The sequence is otherwise terminated at the ith step if there exists an x_j, $j < i$, such that $x_j = x_i$.

Elements of a set $\{x_i'\}$, where $x_i' = 2D - x_i$, are then also discontinuity arguments, because of symmetry. The fact that $\{x_i\}$ and $\{x_i'\}$ include all discontinuity arguments is evident from considerations similar to those used to prove no jumps are positive: in general, any discontinuity other than the pair resulting from discontinuities in the defining equation (at x = 1; x = 2D - 1) implies the existence of a predecessor (in the sense of the sequence defined by (5.2)) at which the jump is twice as large, because of the coefficient,

[10] $x_i = \frac{1}{2}$ and $x_i = D$ are limiting cases where two otherwise distinct discontinuities become coincident.

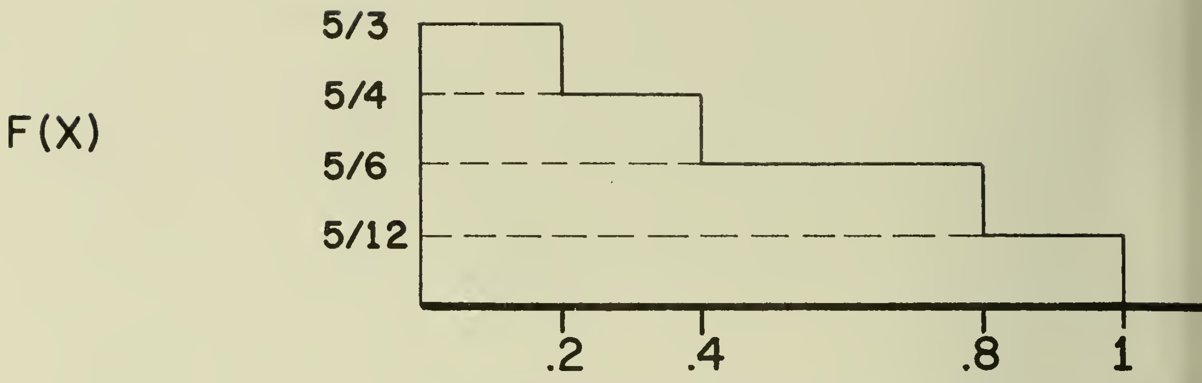


FIGURE 4a: SOLUTION FOR D=3/5

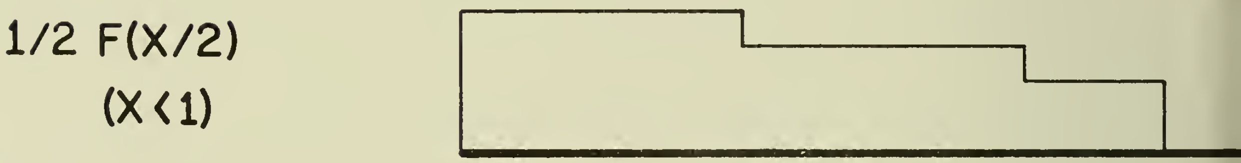


FIGURE 4b

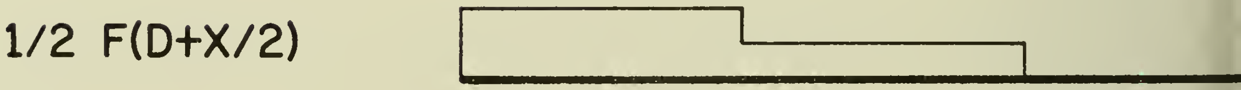


FIGURE 4c

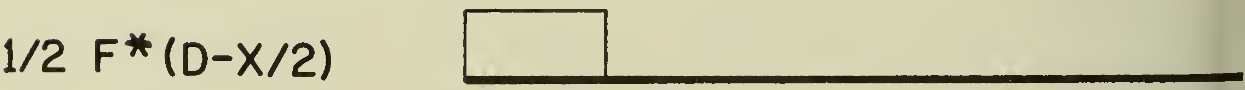


FIGURE 4d

$\frac{1}{2}$, on the right in (1.5).[11] Since the two just cited are the only x_i which do not require such a predecessor, all discontinuities must be descendent from these, in the sense of (5.2)

With the number of discontinuities known, the size of the respective jumps can also be easily evaluated if we ignore the set of divisors, $\{D_E\}$. Because F(x) is monotonic, the sum of the jumps for $x > 0$ must be $\frac{1}{D}$ (see (4.2)). Moreover, the sizes of the jumps in the sequence $\{x_i\}$ form a geometric progression, i.e., the jump at $x = x_{i+1}$ is twice as large as the jump at $x = x_{i-1}$ because of the coefficient of $\frac{1}{2}$ in equation (1.5). Denote

$$a_i \equiv F(x_i-) - F(x_i+) \tag{5.3}$$

then for T discontinuity pairs:

$$\frac{1}{D} = 2 \sum_{i=1}^{T} a_i = 2 \sum_{i=0}^{T-1} a_1 (\tfrac{1}{2})^i = 4a_1(1 - (\tfrac{1}{2})^T) \tag{5.4}$$

or

$$a_1 = \frac{2^{T-2}}{(2^T - 1)} \tag{5.5}$$

(The coefficient, 2, in (5.4) accounts for the fact that the $\{a_i\}$ are half of the jumps.) Therefore, the jump at x_i is:

$$a_i = (\tfrac{1}{2})^{i-1} \circ a_1 = \frac{2^{T-i-1}}{D(2^T - 1)} \tag{5.6}$$

[11] If the divisor is in $\{D_E\}$, or either of the cases $x_i = D$ or $x_i = \frac{1}{2}$ occur, the predecessor requirement applies to all except one "generation" of discontinuities, the exception resulting from the fact that one jump is half the <u>sum</u> of two others.

Jumps at corresponding symmetric discontinuities are, of course, equal.

The solutions derived using (5.2) and (5.6) apply to intervals of the divisor range, rather than individual values of the divisor (excluding solutions for $\{D_E\}$), since there are T pairs of discontinuities in F(x) for all divisors which result in x_T being included in the interval $(\frac{1}{2}, 2D - \frac{1}{2})$. As demonstrated below, there are an infinite number of such intervals, each of which involves a distinct sequence of operations when applying (5.2), though several non-adjacent intervals may result in the same number of discontinuities. Equation (5.1) gives the solution for one of these intervals.

To formulate systematically the R_i distributions for $\frac{1}{2} < D < \frac{3}{4}$, consider the intervals of D for which $(\frac{1}{2})^{n+1} < (2D - 1) < (\frac{1}{2})^n$, n = 1, 2, ..., is true; denote these intervals of D (viz.: $\frac{1}{2} + (\frac{1}{2})^{n+2} < D < \frac{1}{2} + (\frac{1}{2})^{n+1}$, n = 1, 2, ...) as I_n, n = 1, 2, ..., respectively. To express a sequence of discontinuities generated by use of (5.2), the following definitions will be useful:

N(k) ≡ k consecutive normalizing operations (i.e., (5.2a))

S ≡ one application of (5.2b)

(Listing the operations required to form the discontinuity arguments, rather than listing the arguments themselves, will enable observing similarities in different intervals of D which can then be used as a basis to form infinite families of solution intervals.)

Then every I_n includes a subinterval for which the sequence of operations in (5.2) is N(n), i.e., the first $x_i > \frac{1}{2}$ also satisfies (5.2c). In general, I_n will also include a subinterval for which the discontinuity locations in F(x) are determined in (5.2) by the sequence N(n)S; another for which the sequence is N(n)SN; etc. Table 1 is a list of all subintervals of I_n, $n \geq 2$, for which the divisor results in n + 5 discontinuity pairs or less (excluding $\{D_E\}$). The following two examples are provided to show how Table 1 was derived; the solution for I_1 appears in the first example.

Example 1. Consider the subinterval of I_n for which the sequence in (5.2) is N(n); this means the only discontinuities in F(x), for all values of D in this subinterval are at: $x_1 = 2D - 1$; $x_2 = 2(2D - 1)$; ...; $x_n = 2^n(2D - 1)$, together with the respective 2D complements. For this portion of I_n, the following is true:

$$\frac{1}{2} \leq 2^n(2D - 1) < 2D - \frac{1}{2} \qquad (5.7)$$

or

$$\frac{2^{n+1} + 1}{2^{n+2}} \leq D < \frac{2^{n+1} - 1}{2^{n+2} - 4}, \qquad n = 1, 2, \ldots \qquad (5.8)$$

(See Figure 5.) Note that I_1, $.625 \leq D < .75$, is covered completely in this family of solutions. Since the number of discontinuity pairs, T, is two for I_1 (in general, T exceeds the number of steps in (5.2) by one), we find by use of (5.6):

TABLE 1. SOLUTION INTERVALS IN THE DIVISOR RANGE

2:(5.2) Sequence	Interval Bounds	Examples (to eight significant figures)			
		n = 2	n = 3	n = 4	n = 5
1) N(n)	$\frac{2^{n+1}-1}{2^{n+1}-4}$	.583$\underline{3}$....	.53571428	.51$\underline{6}$.....	.50806452
	$\frac{2^{n+1}+1}{2^{n+2}}$	.5625	.53125	.515625	.5078125
2) N(n)S	$\frac{2^{n+2}-1}{2^{n+3}-8}$	.625	.55357143	.525	.51209677
	$\frac{2^{n+2}+1}{2^{n+3}-4}$	.60714286	.55	.52419355	.51190476
3a) N(n)SN	$\frac{2^{n+3}-1}{2^{n+4}-12}$	.59613846	.54310345	.52049180	.51
	$\frac{2^{n+3}-1}{2^{n+4}-8}$	.58928571	.541$\underline{6}$....	.52016129	.50992603
3b) N(n)SS	$\frac{2^{n+3}-1}{2^{n+4}-16}$	....	.5625	.5291$\underline{6}$...	.51411290
	$\frac{2^{n+3}+1}{2^{n+4}-12}$	....	.56034483	.52868852	.514
4a) N(n)SNN	$\frac{2^{n+4}-1}{2^{n+5}-20}$	....	.52813559	.51829263	.50896414
	$\frac{2^{n+4}+1}{2^{n+5}-16}$	....	.5375	.51814516	.50892857
4b) N(n)SNS	$\frac{2^{n+4}-1}{2^{n+5}-24}$	.60576923	.54741379	.52254098	.511
	$\frac{2^{n+4}+1}{2^{n+5}-20}$	.60185185	.54661017	.52235772	.51095618
4c) N(n)SSN	$\frac{2^{n+4}-1}{2^{n+5}-28}$	....	.55701754	.52685950	.51305221
	$\frac{2^{n+4}+1}{2^{n+5}-24}$	....	.55603448	.52663934	.513
4d) N(n)SSS	$\frac{2^{n+4}-1}{2^{n+5}-32}$	....	....	.53125	.51512097
	$\frac{2^{n+4}+1}{2^{n+5}-28}$	....	....	.53099174	.51506024

TABLE 1. SOLUTION INTERVALS IN THE DIVISOR RANGE (CONT'D)

2:(5.2) Sequence	Interval Bounds	Examples (to eight significant figures)			
		n = 2	n = 3	n = 4	n = 5
5a) N(n)SNNN	$\frac{2^{n+5}-1}{2^{n+6}-36}$	....	....	.51720648	.50844930
	$\frac{2^{n+5}+1}{2^{n+6}-32}$	....	....	.51713710	.50843254
5b) N(n)SNNS	$\frac{2^{n+5}-1}{2^{n+6}-40}$	.58796296	.54025424	.51930894	.50946215
	$\frac{2^{n+5}+1}{2^{n+6}-36}$	.58636364	.53991597	.51923077	.50944334
5c) N(n)SNSN	$\frac{2^{n+5}-1}{2^{n+6}-44}$	.59905660	.54487179	.52142857	.51047904
	$\frac{2^{n+5}+1}{2^{n+6}-40}$	.597$\underline{2}$....	.54449153	.52134146	.51045817
5d) N(n)SNSS	$\frac{2^{n+5}-1}{2^{n+6}-48}$	....	.54956897	.52356557	.5115
	$\frac{2^{n+5}+1}{2^{n+6}-44}$	....	.54914530	.52346939	.51147705
5e) N(n)SSNN	$\frac{2^{n+5}-1}{2^{n+6}-52}$	....	.55434783	.52572016	.51252505
	$\frac{2^{n+5}+1}{2^{n+6}-48}$	....	.55387931	.52561475	.5125
5f) N(n)SSNS	$\frac{2^{n+5}-1}{2^{n+6}-56}$	....	.55921053	.52789256	.51355422
	$\frac{2^{n+5}+1}{2^{n+6}-52}$	....	.55869565	.527$\underline{7}$....	.51352705
5g) N(n)SSSN	$\frac{2^{n+5}-1}{2^{n+6}-60}$	....	....	.53008299	.51458752
	$\frac{2^{n+5}+1}{2^{n+6}-56}$	....	....	.52995868	.51455823
5h) N(n)SSSS	$\frac{2^{n+5}-1}{2^{n+6}-64}$	....	....	....	.515625
	$\frac{2^{n+5}+1}{2^{n+6}-60}$	....	....	....	.51559356

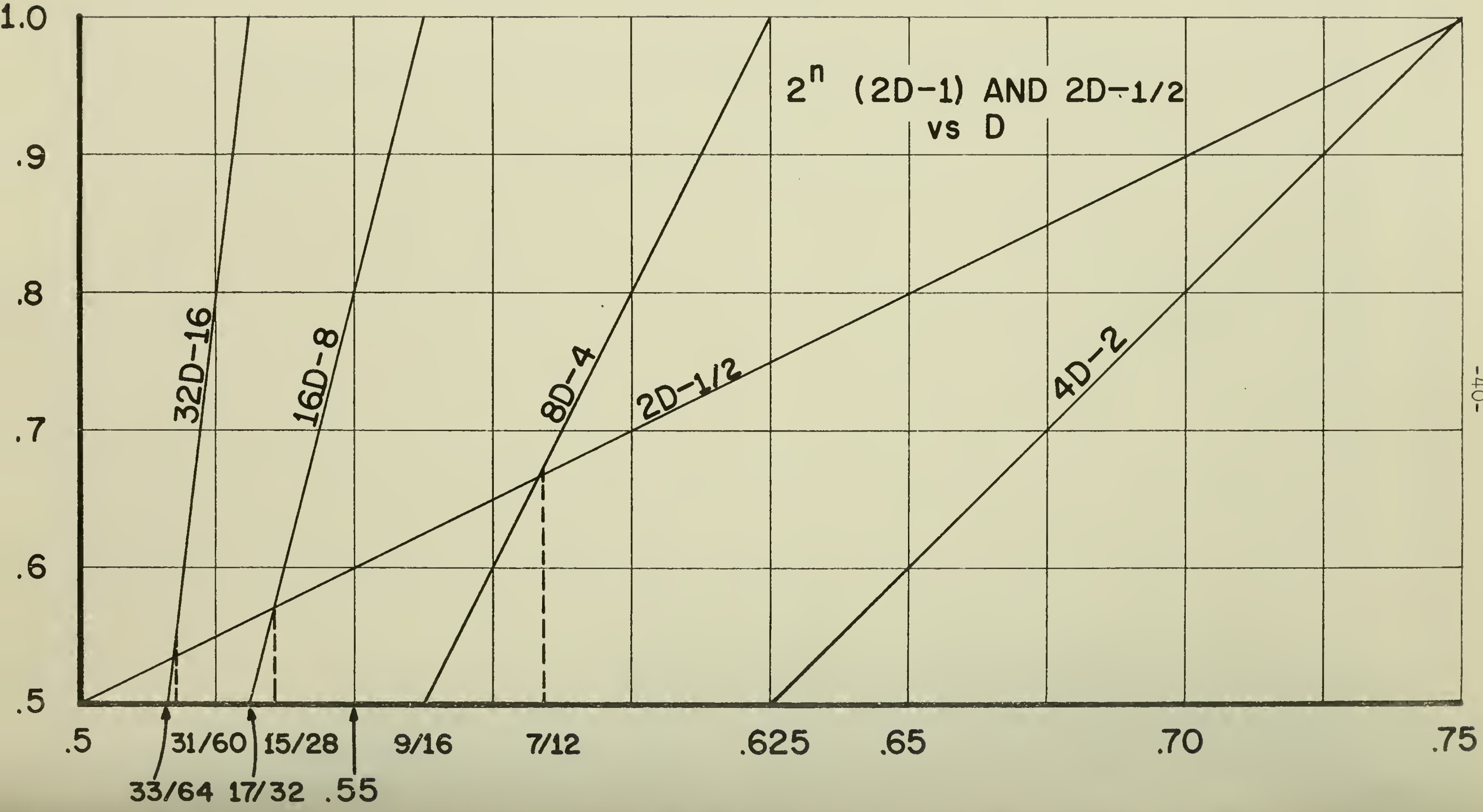


FIGURE 5: INTERVALS FOR WHICH (5.2) SEQUENCE

$$F(x) = \begin{cases} \frac{1}{D}, & 0 < x < 2D - 1 \\ \frac{2}{3D}, & 2D - 1 < x < \min(4D - 2, 2 - 2D) \\ \frac{1}{2D}, & \min(4D - 2, 2 - 2D) < x < \max(4D - 2, 2 - 2D) \\ \frac{1}{3D}, & \max(4D - 2, 2 - 2D) < x < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (5.9)$$

For D in any given solution subinterval, x_T and $x_T' = 2D - x_T$ (and <u>only</u> these two discontinuity arguments) may be related by any of: $x_T > x_T'$, $x_T' > x_T$, or $x_T' = x_T = D$. This is the reason for the "max" and "min" in (5.9).

Example 2. If the sequence in (5.2) is N(n)SS, the following relation is satisfied by D:

$$\frac{1}{2} < 2\{2^n(2D - 1) - D] - D\} < 2D - \frac{1}{2} \qquad (5.10)$$

or

$$\frac{2^{n+3} + 1}{2^{n+4} - 12} < D < \frac{2^{n+3} - 1}{2^{n+4} - 16} \qquad (5.11)$$

There are two reasons why any given sequence of operations in (5.2) may not correspond to a solution interval in the range of D:

a) Some intermediate term may necessarily satisfy $\frac{1}{2} < x_i < 2D - \frac{1}{2}$, if the implied final term also does.

b) The sequence may not be self-consistent; e.g., if a sequence begins with N(k)S ..., then $(2D - 1) > (\frac{1}{2})^{k+1}$

is true, and therefore an inconsistency would exist if $k + 1$ (or more) successive normalizations occurred at any point subsequently, because $2D - 1$ is the smallest discontinuity argument.

In the example being considered, N(1)SS cannot correspond to a solution because all values of D for which only one step is required to normalize $2D - 1$ are part of the interval in which N(1) is the complete sequence. N(2)SS cannot yield a solution since, if the first three steps are N(2)S, it can easily be shown that $x_4 < 2D - \frac{1}{2}$, and therefore the sequence must either terminate after N(2)S, or be followed by a normalization as the next step.

For all $n \geq 3$, equation (5.11) correctly defines intervals of D for which the discontinuities in F(x) are generated by N(n)SS in (5.2), since neither of the two reasons for exclusion are violated in these areas.

To facilitate deciding whether a given sequence of operations when applying (5.2) corresponds to some interval of D, there are four simple rules to test for consistency.

1) In a sequence starting with N(k)S, the maximum number of successive normalizations which can occur anywhere in the sequence is k. This is because $2D - 1$ is the smallest discontinuity argument.

2) A sequence starting with N(k)S cannot also terminate with N(k). Consider Figure 5 for a given value of D. If $2^k(2D - 1) > 2D - \frac{1}{2}$, then any other x_i occurring in the $(\frac{1}{2})^{k+1} < x < (\frac{1}{2})^k$ interval will, after normalization, also exceed $2D - \frac{1}{2}$, again because all $x_i \geq 2D - 1$.

The other two restrictions are simply duals of the above. Recalling that (5.2) is generating half of the total set of discontinuity arguments, it may be noted the sequence of operations to generate the 2D complement set,

$\{x_i'\}$, starting with $x_i' = 1$, is the dual (i.e., interchange N's and S's) of the sequence to generate $\{x_i\}$. To determine the bound on the number of successive S's which may occur for D in a given interval, I_m, we assume D is such that some x_j is almost equal to one. Rather than examining bounds on x_{j+1}, etc., attention may, at that point, be shifted to the 2D complement sequence x_j', x_{j+1}', etc., and the limits on the number of successive normalizations in the complement sequence determined. From these considerations one obtains:

3) A sequence starting with N(k)S can have at most k successive S's.

4) A sequence starting with N(k)S cannot terminate with a subsequence of k successive S's.

After rejecting by inspection those sequences which are not consistent, the final test for the validity of a proposed sequence of operations in (5.2) is comparison of the interval of D implied with intervals corresponding to all subsequences (formed by truncation) to determine whether an intermediate x_i was in the $(\frac{1}{2}, 2D - \frac{1}{2})$ interval as well as the implied final one.

It is therefore only necessary to find the smallest value of n, say n_0, for which a given sequence in Table 1 correctly describes a solution interval, and the expression then applies for all $n \geq n_0$. Solutions in Table 1 are arranged in the order of increasing complexity, such that all subintervals of I_n for which F(x) has k discontinuity pairs are listed before those with k + 1. Moreover, by ordering those which have the same number of discontinuities for I_n according to a binary weighting, such that S corresponds to 1 and N to zero, it happens that constants in the denominators of successive upper bounds (as well as successive lower bounds) form a readily discernible arithmetic progression. If desired, this progression may be used to extend the table indefinitely.

2.5.1 Properties of Solution Intervals

Though by no means unique, the ordering of solution intervals given by Table 1 displays one of several interesting characteristics of the solutions. Aside from the finite number of I_n for which a given form of sequence does not apply, all I_n include:

a) one subinterval which yields n + 1 discontinuity pairs in F(x);

b) one for which there are n + 2 pairs;

c) two for which there are n + 3 pairs;

d) four for which there are n + 4 pairs;

and in general,

e) 2^j for which there are n + j + 2 pairs.

I_5, represented by the column for n = 5, is an example of this, up to ten discontinuity pairs.

A second property which is apparent from Table 1 is that solution intervals are of the form

$$\frac{2^m + 1}{2^{m+1} - B} < D < \frac{2^m - 1}{2^{m+1} - (B + 4)} \qquad (5.12)$$

where B and m are integers. This can be proved by observing that for an arbitrary sequence of operations in (5.2), the inequalities which determine the solution interval (e.g., (5.07), (5.10)) are of the form:

$$\frac{1}{2} < 2^j(2D - 1) - kD < 2D - \frac{1}{2} \qquad (5.13)$$

where j and k are integers. (5.13) then yields an expression of the form of (5.12).

Consider now the problem of finding contiguous solution intervals, rather than the successively complex but usually nonadjacent intervals of Table 1. Let $d_1 < D < d_2$ represent a solution interval for which the discontinuity arguments generated by (5.2) are y_1, y_2, ..., y_T, each expressed as a function of D. For D in the region immediately to the right of this interval, y_T exceeds $2D - \frac{1}{2}$, by definition of an interval boundary (see (5.2c)). The only difference in the first T discontinuity arguments for $D = d_2+$, as compared with those for $d_1 < D < d_2$, would be if some y_j, $j < T$, satisfied $\frac{1}{2} < y_j < 2D - \frac{1}{2}$, resulting in an interval of fewer discontinuitites. Otherwise stated, since all y_i are of the form bD - C, where b and c are positive, the only change in the relations of the y_i to the thresholds in (5.2a) and (5.2b) which can occur by increasing D is that one (or more) of the y_i cross $\frac{1}{2}$. If, however, such a y_j, $j < T$ does not exceed $\frac{1}{2}$, then d_2 is an accumulation point of solution intervals! To prove this, we note that for $D = d_2 + \epsilon$, and ϵ small but positive:

$$y_T = (2D - \frac{1}{2}) + p \circ 2^T\epsilon$$

where $\frac{1}{2} < p \leq 1$, and therefore

$$y_{T+1} = 2[y_T - D] = 2D - 1 + p \circ 2^{T+1}\epsilon$$

For ϵ sufficiently small, the process defined in (5.2) can be made to approximate an infinite cycling for an arbitrary number of iterations since y_{T+1} approximates y_1, and the perturbation introduced by ϵ causes the term which corresponds to y_T in each iteration to move further from the $2D - \frac{1}{2}$ threshold.

To summarize, the above considerations reveal any solution interval is bounded on the right by either:

a) an accumulation point of solution intervals, or

b) a solution interval yielding fewer discontinuities.

There is never a single adjacent interval on the right which yields a greater number of discontinuitites.

Next consider the relation of solution interval $d_1 < D < d_2$ to the adjacent interval on the left, again denoting the terms generated by (5.2) for $d_1 < D < d_2$ as $y_1, y_2, \ldots, y_T$. Due to the definition of an interval boundary, as D is varied from d_1+ to d_1-, the y_T term moves out of the $(\frac{1}{2}, 2D - \frac{1}{2})$ interval; moreover, none of the y_j, $j < T$ replace y_T in that interval since this would imply that $d_1 < D < d_2$ is bordered on the left with an interval of fewer discontinuity pairs, which contradicts what was just established above. Therefore, the inequality relation of each y_j, $j < T$, with respect to the $(\frac{1}{2}, 2D - \frac{1}{2})$ interval is not altered as D changes from d_1+ to d_1-, which means the first T terms generated in (5.2) for $D = d_1-$ are the same as those for $d_1 < D < d_2$. If $D = d_1 - \epsilon$, with ϵ small, then $y_T = \frac{1}{2} - p \cdot 2^T \cdot \epsilon$, where $\frac{1}{2} < p \leq 1$; $y_{T+1} = 1 - p2^T \cdot \epsilon$; $y_{T+2} = 2 - 2D - p \cdot 2^{T+1}\epsilon$; .... Starting with y_{T+1}, these terms may be expressed relative to the 2D complements of the first T terms as follows: $y_{T+1} = y_1' - p \cdot 2^T\epsilon$, $y_{T+2} = y_2' - p \cdot 2^{T+2}\epsilon$, ..., $y_{2T} = y_T' - p \cdot 2^{2T}\epsilon$. Now, since y_T is $p \cdot 2^T\epsilon$ away from the $(\frac{1}{2}, 2D - \frac{1}{2})$ interval, y_T' is as well, due to symmetry, and therefore these expressions reveal that y_{2T} must be inside the interval. Therefore, $d_1 < D < d_2$, the generic solution interval, is bounded on the left by an interval with twice the number of discontinuity pairs.

By induction, the results of the last three paragraphs yield the

<u>Theorem 2.5c</u>: Every solution interval is an element of an infinite sequence adjacent intervals, each of which results in half the numbers of discontinuity pairs in F(x) as the interval at its left. Except for the sequence beginning with the $\frac{3}{4} < D < 1$ interval, the first interval in each such sequence is bordered on the right by an accumulation point of solution intervals.

An example of such an infinite sequence begins with $\frac{3}{4} < D < 1$, for which F(x) has one discontinuity pair. The next four intervals are:

a) $\frac{5}{8} < D < \frac{3}{4}$

b) $\frac{17}{28} < D < \frac{15}{24}$

c) $\frac{257}{424} < D < \frac{255}{420}$

and

d) $\frac{65537}{108124} < D < \frac{65535}{108120}$, in which there are 2, 4, 8, and 16 discontinuity pairs, respectively.

The discontinuity arguments which satisfy the $\frac{1}{2} < x_i < 2D - \frac{1}{2}$ relation for the last three examples are shown in Figure 6, which illustrates the rapid convergence of such a sequence. It happens that all the solution intervals appearing in Table 1, except those described by N(2)S, N(3)SS, N(4)SSS, and N(5)SSSS, are examples of first intervals in infinite sequences. Due to the above theorem, it is only necessary to observe whether any of the intervals resulting in fewer discontinuitites than the interval in question is adjacent, to decide whether a given interval is first in such a sequence.

Parenthetically, it may be noted that equation (5.12) provides a very straightforward method for computing boundaries on successive adjacent intervals in these infinite sequences. As an example, $\frac{17}{28} = \frac{2^4 + 1}{28}$ is of the form of a lower bound in (5.12). To express the same number in the form of an upper bound, $\frac{17}{28} \cdot \frac{2^4 - 1}{2^4 - 1} = \frac{2^8 - 1}{420}$. Then the lower bound on the same interval is $\frac{255 + 2}{420 + 4}$, again using (5.12). In addition, the list of operations when applying (5.2) for each interval of D in such a sequence may be generated by application of the following rule:

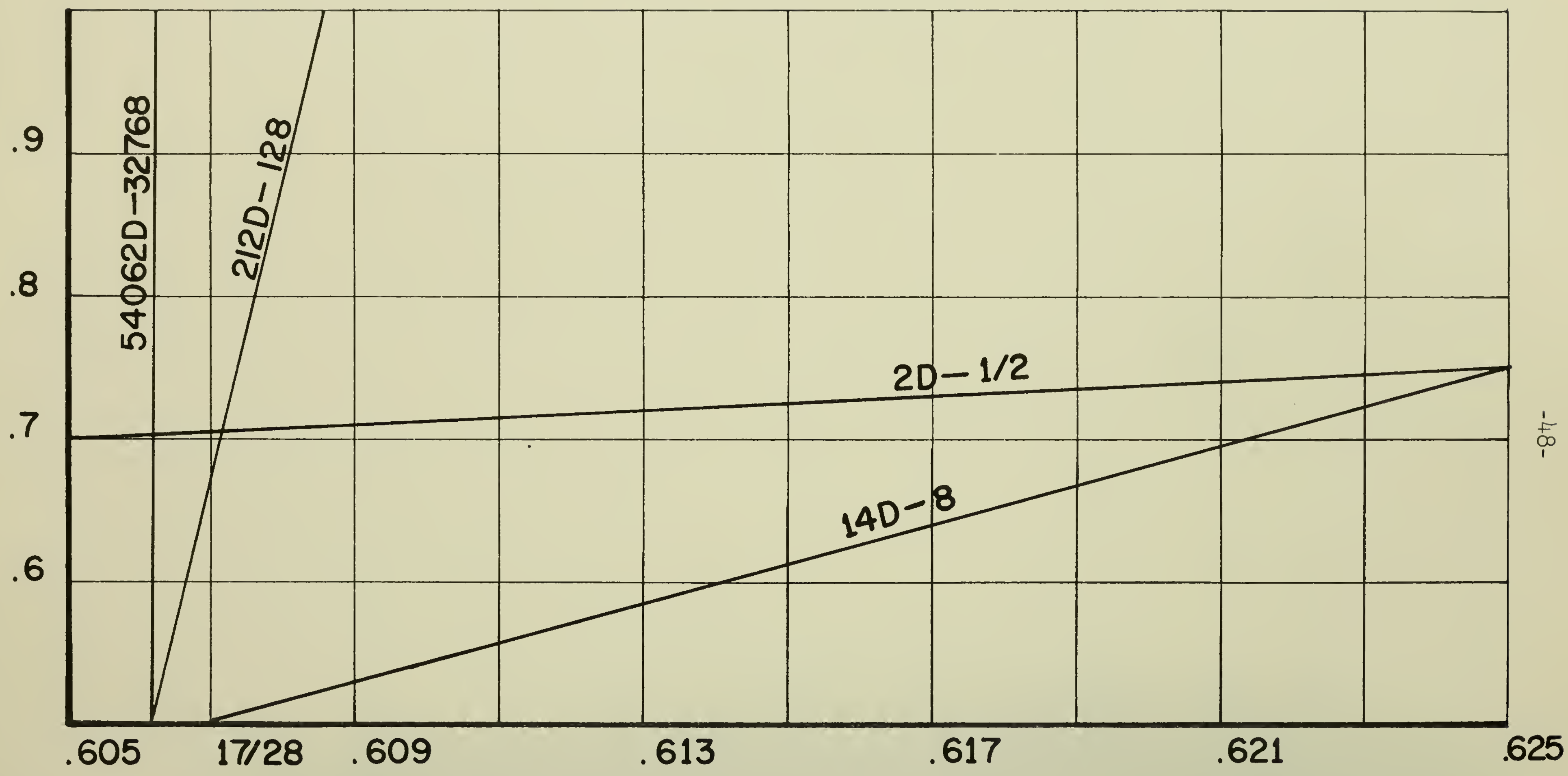


FIGURE 6: X_T vs D FOR ONE GROUP OF CONTIGUOUS INTERVALS.

If $u(N,S)$ denotes the N-S sequence corresponding to any given solution interval, and $\bar{u}(N,S)$ is defined to be the same sequence with N's and S's interchanged, then the solution interval to the left of the one in question is characterized by $u(N,S)N\bar{u}(N,S)$. (5.14)

This rule is merely a restatement of an observation which led to Theorem 2.5c, where it was noted that with K discontinuity pairs, the second $\frac{K}{2}$ terms generated by use of (5.2) closely approximate the 2D complements of the first $\frac{K}{2}$. Applying it to the four intervals mentioned above, one obtains:

$\frac{5}{8} < D < \frac{3}{4}$:	N
$\frac{17}{28} < D < \frac{5}{8}$:	N N S
$\frac{257}{424} < D < \frac{17}{28}$:	NNS N SSN
$\frac{65337}{108124} < D < \frac{65535}{108120}$:	NNSNSSN N SSNSNNS

As a final observation, an immediate corollary to Theorem 2.5c is that all solution intervals with odd numbers of discontinuity pairs have an accumulation point of discontinuitites as an upper bound, i.e., such intervals are first in an infinite sequence of the type described by Theorem 2.5c.

2.5.2 Comments on the Set $\{D_E\}$

The exceptional divisors, $\{D_E\}$, defined to be those values of D for which (5.2d) is satisfied, may be recognized as one of three distinctly different types of limit points (or accumulation points) which would exist if Table 1 were extended indefinitely. One type of limit point is the right edge of certain solution intervals; for D arbitrarily close but greater than such a boundary, the sequence of operations in (5.2) appears to cycle as discussed in the previous section. However, the $\{D_E\}$ are accumulation points of solution intervals on either side, unlike these interval boundary points. The third type of limit point is that to which an infinite sequence of contiguous intervals converges (see Theorem 2.5c). This type of accumulation point of solution intervals is distinguished from the other two by the fact that the sequence of operations when applying (5.2) does not cycle. See (5.14) and the examples which follow it.

To generate a few subsets of $\{D_E\}$, consider the following examples of cycling which could arise when applying (5.2):

a) $x_1 = 2D - 1$; $x_2 - 4D - 2$; $x_3 - 8D - 4$;
$x_4 = 14D - 8 = 4D - 2$, i.e., $D = \frac{3}{5}$.

b) $x_1 = 2D - 1$; $x_2 = 4D - 2$; $x_3 = 8D - 4$; $x_4 = 16D - 8$;
$x_5 = 30D - 16 = 40 - 2$, i.e., $D = \frac{7}{13}$.

The first of these is an example of the subset of $\{D_E\}$ for which the sequence of operations is $N(n)SN\underline{SN}$ ...; n = 2, 3, 4, ... The second is an example of $N(n)SNN\underline{SNN}$ ...; n = 3, 4, ... The values of D in the first of these subsets satisfy the equation

$$2^n(2D - 1) = 2[2^{n+1}(2D - 1) - D], \quad n = 1, 2, \ldots \qquad (5.14)$$

and therefore the subset is:

$$D = \frac{3(2^n)}{3 \cdot 2^{n+1} - 2}; \quad n = 1, 2, \ldots \qquad (5.15)$$

The values of D in the second subset are described by:

$$2^n(2D - 1) = 2[2^{n+2}(2D - 1) - D] \qquad (5.16)$$

or

$$D = \frac{7(2^n)}{7 \cdot 2^{n+1} - 2}, \qquad n = 1, 2, \ldots \qquad (5.17)$$

Both (5.16) and (5.17) can be shown to be proper subsets of the set in $\{D_E\}$:

$$D = \frac{n}{2^n - 1},$$

where n assumes all integer values excluding powers of two.

3. SHIFT AVERAGE

3.1 Equation for <S>

While the density of partial remainders, as a function of D, is interesting in itself, its practical value lies in evaluating what fraction of the total number of quotient digits produced are nonzero. This figure of merit is frequently expressed in reciprocal form as the shift average, i.e., the average number of binary shifts in the division process between successive uses (addition or subtraction) of the divisor. The shift occurring as part of a step producing a nonzero quotient digit is counted when computing the shift average. Define:

$S(D) = S \equiv$ random variable denoting the shift count between successive uses of divisor D;

$\langle S \rangle$ = expectation value of S.

Since the distribution of R_i is independent of the dividend (for i sufficiently large), the distribution of S is also. This independence of R_i was based on the assumption that the probability density describing dividends is piecewise constant, and finite (Section 1.3.2). Therefore, for a divisor D_0, $\langle S(D_0) \rangle = C$ means that the shift average for any ensemble of dividends which satisfies the above assumption is C; it does not mean that with divisor equal to D_0, the shift average is C for each individual dividend.

If the distribution of the shift count, S, were known, <S> could be evaluated by use of the relation which defines expectation:

$$\langle S \rangle = \sum_{m=1}^{\infty} m \cdot \Pr[S = m] \qquad (1.1)$$

$\Pr[S = m]$ can be determined by converting $f(x)$, the distribution of R_i to the distribution of normalized partial remainders, $g(x)$, using equation 2:(1.7); then:

$$\Pr[S = m] = g(D + \frac{1}{2^m}) - g(D + \frac{1}{2^{m+1}}) + g(D - \frac{1}{2^{m+1}}) - g(D - \frac{1}{2^m}) \tag{1.2}$$

Alternatively, $<S>$ can be determined without knowledge of the distribution of S if we make use of a rather basic property of ergodic, aperiodic Markov processes:[1]

$$\begin{aligned} &\text{if} \quad M_j \equiv \text{mean recurrence time of state } j, \\ &\text{and} \quad \pi_j \equiv \text{absolute probability of state } j, \\ &\text{then} \quad M_j = (\pi_j)^{-1}. \end{aligned} \tag{1.3}$$

Since S can be regarded as the recurrence time of the set of states for which $x > \frac{1}{2}$ (where steps of the division process correspond to units of time in the terminology of a Markov process), equation (1.3) implies:

$$<S>^{-1} = \Pr[R_i > \tfrac{1}{2}]$$

or

$$<S> = \frac{1}{1 - f(\frac{1}{2})}$$

Therefore, to find the shift average, it is only necessary to find the area under the density function, $F(x)$, for $x > \frac{1}{2}$ (or $x < \frac{1}{2}$, whichever is

[1] e.g., see Feller, Chapt. XV, Section 5.

more convenient). This simplification in evaluating $\langle S\rangle$ is a by-product of considering the distribution of R_i, rather than that of normalized partial remainders.

3.2 Evaluation of $\langle S(D)\rangle$ for $\frac{3}{5} < D < 1$

F(x), for $\frac{3}{4} < D < 1$, is given by equation 2:(5.1), and we can easily confirm Freiman's result for this region by use of (1.4):

$$\langle S\rangle = \frac{1}{1 - \frac{1}{2D}} = \frac{2D}{2D - 1}; \quad \frac{3}{4} < D < 1 \tag{2.1}$$

For $\frac{5}{8} < D < \frac{3}{4}$, F(x) is given by equation 2:(5.9), from which we obtain:

$$f(\tfrac{1}{2}) = (2D - 1)\frac{1}{D} + \frac{2}{3D}(\frac{1}{2} - 2D + 1)$$

$$= \frac{2}{3}, \text{ or} \tag{2.2}$$

$$\langle S\rangle = 3, \text{ for } \frac{5}{8} < D < \frac{3}{4}$$

The next adjacent interval in the range of D is $\frac{17}{28} < D < \frac{5}{8}$, which appears in Table 1 as the n = 2 entry for sequence N(n)S. Since the N(2)S sequence involves three steps, there are four discontinuity pairs in F(x): 2D - 1, 4D - 2, 8D - 4, and 14D - 8, together with the respective 2D complements. The area under F(x), $x < \frac{1}{2}$, can be determined by use of equation 2:(5.6). First:

$$F(\tfrac{1}{2}) = F(D) + a_T$$

$$= \frac{1}{2D} + \frac{1}{2D}(\frac{1}{2^T - 1}) = \frac{1}{2D}(\frac{2^T}{2^T - 1}) \tag{2.3}$$

where, as defined in Chapter 2, T is the number of discontinuity pairs, and a_T is the jump in F(x) at one of the two discontinuities in the $(\frac{1}{2}, 2D - \frac{1}{2})$ interval, provided D is not in the set $\{D_E\}$. The area in question is then expressible as:

$$f(\tfrac{1}{2}) = \tfrac{1}{2} \cdot F(\tfrac{1}{2}) + (2D - 1)a_1 + (4D - 2)a_2 + (4 - 6D)a_3$$

$$= \frac{2^{T-2}}{D(2^T - 1)} [1 + 2D - 1 + \frac{4D - 2}{2} + \frac{4 - 6D}{4}] \qquad (2.4)$$

$$= \frac{2}{3}$$

(Note that 2D - 1, 4D - 2, and 4 - 6D are the only discontinuity arguments less than $\frac{1}{2}$ for D in this interval.) Therefore from (1.4):

$$\langle S \rangle = 3, \quad \text{for } \frac{17}{28} < D < \frac{5}{8} \qquad (2.5)$$

Thus far, the shift averages for all divisors in the range $\frac{17}{28} < D < 1$ have been found. In addition, it can be readily ascertained from Figure 4, using equation (1.4), that the shift average for the point D = .6 is also 3. The conjecture that $\langle S \rangle = 3$ holds for the interval $.6 < D < \frac{17}{28}$ is a natural consequence of these observations. However, the number of solution intervals in this small interval of D is infinite, with the number of discontinuities in F(x) becoming arbitrarily large in an infinite number of distinct neighborhoods.

The complex behavior of F(x) for $\frac{3}{5} < D < \frac{5}{8}$ notwithstanding, the conjecture raised regarding $\langle S \rangle$ in this region can be proved by use of the symmetry property, derived in Section 2.3, as demonstrated by the following two theorems.

Theorem 3.2a: There are no discontinuities in F(x) in the interval $4 - 6D < x < 4D - 2$, for $\frac{3}{5} < D < \frac{5}{8}$.

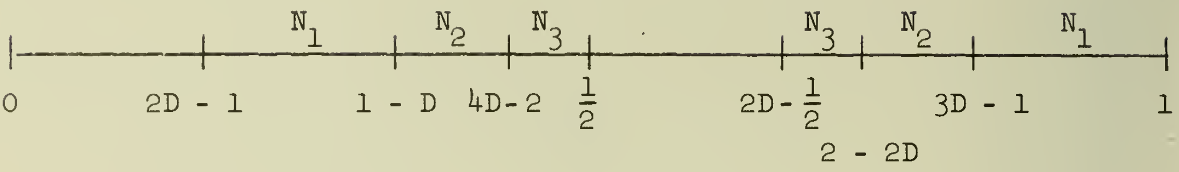


FIGURE 7

Proof:

1. Let N_1, N_2, and N_3 denote the numbers of discontinuity arguments appearing in the intervals $2D - 1 < x \leq 1 - D$, $1 - D < x < 4D - 2$, and $4D - 2 \leq x < \frac{1}{2}$, respectively. Due to symmetry, the same numbers of discontinuities, respectively, appear in $3D - 1 \leq x < 1$, $2 - 2D < x < 3D - 1$, and $2D - \frac{1}{2} < x \leq 2 - 2D$. See Figure 7. (Note that the ordering of the end points of these intervals with respect to size remains as shown in Figure 7 for the interval of D in question.)
2. For $D < \frac{3}{4}$, and $2 - 2D < x < 1$, equation 2:(1.5) reduces to:

$$F(x) = \frac{1}{2} F(\frac{x}{2}) \tag{2.6}$$

In words, F(x) for $2 - 2D < x < 1$ is a copy of F(x) for $1 - D < x < .5$ (though with horizontal scale doubled and vertical scale halved); therefore:

$$N_2 + N_3 = N_2 + N_1$$

or (2.7)

$$N_3 = N_1$$

3. Due to equation (2.6), all N_3 discontinuities in $4D - 2 \leq x < \frac{1}{2}$ map into $8D - 4 \leq x < 1$. This, together with equation (2.7), implies no discontinuities in $1 - D < x < 4D - 2$ map into $3D - 1 < x < 1$.

4. Therefore, the N_2 discontinuities in $1 - D < x < 4D - 2$ are related to those in $2 - 2D < x < 3D - 1$ in two respects: symmetry, and mapping via equation (2.6). Due to the latter, if x_K denotes a discontinuity argument such that $1 - D < x_K < 4D - 2$, the jump at $2x_K$ is half as large. Attempting to select the smallest (nonzero) jump among the N_2 in $1 - D < x < 4D - 2$ would therefore lead to a contradiction, since another jump, half as large, would exist in $2 - 2D < x < 3D - 1$, and also in the first interval due to symmetry. Hence, $N_2 = 0$.

5. In step 3, the upper limit on one of these symmetric discontinuity-free intervals was extended from $3D - 1$ to $8D - 4$. The symmetric mate to $(2 - 2D, 8D - 4)$ is $(4 - 6D, 4D - 2)$. Q.E.D.

Theorem 3.2b: $\langle S \rangle = 3$ for $\frac{3}{5} < D < \frac{5}{8}$.

Proof:

(The proof consists of relating areas under $F(x)$ for certain intervals of x. Recognizing expressions of the form: $f(z_1) - f(z_2)$, as the area under $F(x)$, for $z_2 < x < z_1$ may be found helpful.)

1. Using equation 2:(1.4) for $D < \frac{3}{4}$:

$$f(4D - 2) - f(2D - 1) = f(3D - 1) - f(2D - \tfrac{1}{2}) + f(2D - 1) - f(D - \tfrac{1}{2}) \tag{2.8}$$

2. Since $F(x) = \frac{1}{D}$, for $0 < x < 2D - 1$,

$$f(2D - 1) - f(D - \tfrac{1}{2}) = 1 - \frac{1}{2D} \tag{2.9a}$$

and

$$f(2D - 1) = 2 - \frac{1}{D} \tag{2.9b}$$

3. Substituting (2.9) into (2.8):

$$f(4D - 2) = f(3D - 1) - f(2D - \tfrac{1}{2}) + 3 - \frac{3}{2D} \tag{2.10}$$

4. Due to symmetry, and the fact that $F(D) = \frac{1}{2D}$:

$$F(D + C) + F(D - C) = \frac{1}{D}, \qquad \text{for any } C < D \tag{2.11}$$

Therefore, the sum of areas under $F(x)$ for two symmetrically located intervals, each of width W, is $\frac{W}{D}$, i.e.,

$$\int_{z_1}^{z_2} \{F(D + y) + F(D - y)\}dy = \frac{z_2 - z_1}{D}$$

provided all arguments are in the $(0, 2D)$ interval. An application of this relation is:

$$f(\tfrac{1}{2}) - f(1 - D) + f(3D - 1) - f(2D - \tfrac{1}{2}) = \frac{D - \frac{1}{2}}{D} = 1 - \frac{1}{2D} \tag{2.12}$$

Substituting (2.12) into (2.10):

$$f(4D - 2) = 4 - \frac{2}{D} - [f(\tfrac{1}{2}) - f(1 - D)]$$

or (2.13)

$$f(\tfrac{1}{2}) = 4 - \frac{2}{D} - [f(4D - 2) - f(1 - D)]$$

5. From Theorem 3.2a:

$$F(4D - 2-) = F(1 - D) \tag{2.14}$$

Using 2:(1.5) for $D < \frac{3}{4}$ and $x = 4D - 2-$:

$$\begin{aligned} F(4D - 2-) &= \frac{1}{2}[F(2D - 1-) + F(3D - 1-)] \\ &= \frac{1}{2D} + \frac{1}{2} F(3D - 1-) \end{aligned} \tag{2.15}$$

From equation (2.11):

$$F(3D - 1) + F(1 - D) = \frac{1}{D} \tag{2.16}$$

After substitutions, the last three equations yield:

$$F(4D - 2-) = \frac{2}{3D} \tag{2.17}$$

6. The area under $F(x)$, $1 - D < x < 4D - 2$, can now be evaluated, using (2.14) and (2.17):

$$f(4D - 2) - f(1 - D) = \frac{2}{3D}[4D - 2 - (1 - D)] = \frac{10}{3} - \frac{2}{D} \tag{2.18}$$

Using this result in (2.13), we obtain:

$$f(\tfrac{1}{2}) = \frac{2}{3} \tag{2.19}$$

which produces the desired relation in (1.4). Q.E.D.

The fact that the shift average is equal to 3 for all divisors in $\frac{3}{5} < D < \frac{3}{4}$ was indicated by use of simulation studies by Freiman, though proved only for the subinterval $\frac{17}{28} < D < \frac{3}{4}$. This result has been used by Professor G. Metze in proposing a modification to S-R-T division. Briefly, the Metze form of S-R-T division uses a variable threshold value, rather than the constant $\frac{1}{2}$, to determine whether a given step involves an addition (or subtraction) of the divisor. The threshold is selected at the start of any particular division, based on an inspection of the leading digits of the divisor; if K denotes the threshold, then the relation to be satisfied in order to maximize the shift average is:

$$\frac{6}{5} K < D < \frac{3}{2} K$$

or

$$\frac{2}{3} D < K < \frac{5}{6} D$$

3.3 <u>$\langle S(D) \rangle$ for $\frac{1}{2} \leq D \leq \frac{3}{5}$</u>

For D > .6, $\langle S(D) \rangle$ has been found to be a relatively straightforward function. However, in the interval $\frac{1}{2} < D < .6$, the complex behavior of the partial remainder density is reflected in $\langle S(D) \rangle$ as well.

To evaluate $\langle S(D) \rangle$ for the families of solution intervals for which F(x) was determined in Chapter 2, equation 2:(5.6) will be used to find $f(\frac{1}{2})$. Due to symmetry, together with Theorem 2.5b, one of each pair of discontinuity arguments in less than $\frac{1}{2}$, with the exception of the one pair in $(\frac{1}{2}, 2D - \frac{1}{2})$. If x_i^* denotes which of the pair of symmetric discontinuity arguments, x_i, x_i', is less than $\frac{1}{2}$, then:

$$f(\tfrac{1}{2}) = \tfrac{1}{2} F(\tfrac{1}{2}) + \sum_{i=1}^{T-1} a_i x_i^* \qquad (3.1)$$

where $F(\frac{1}{2})$ is given by equation (2.3). As an example, consider the infinite set of intervals in the range of D for which the sequence of operations in 2:(5.2) was N(n), n = 1, 2, ...; viz.:

$$\frac{2^{n+1}+1}{2^{n+2}} < D < \frac{2^{n+1}-1}{2^{n+2}-4}$$

Using (3.1) and 2:(5.6), and observing that T = n + 1 in this case, we find that:

$$f(\tfrac{1}{2}) = \frac{2^{n-1}}{D(2^{n+1}-1)} [1 + n(2D - 1)], \quad n = 1, 2, \ldots \qquad (3.2)$$

for this family of intervals. As a second example, consider the family of intervals corresponding to the operations N(n)S in 2:(5.2); T = n + 2, and the expression is:

$$f(\tfrac{1}{2}) = \frac{2^{n}}{D(2^{n+2}-1)} [1 + n(2D - 1) + \frac{2D - 2^{n}(2D - 1)}{2^{n}}] \qquad (3.3a)$$

$$= \frac{2^{n}}{D(2^{n+2}-1)} [1 + (n - 1)(2D - 1) + \frac{D}{2^{n-1}}] \qquad (3.3b)$$

In (3.3a), the term n(2D - 1) results from the n multiples of 2D - 1 (i.e., $2^j(2D - 1)$, j = 0, 1, ..., n - 1) which are less than $\frac{1}{2}$. Since the sequence terminates after one subtraction, the discontinuity argument $2^n(2D - 1)$ must exceed $2D - \frac{1}{2}$, and therefore its 2D complement must be less than $\frac{1}{2}$, and is represented by the last term in (3.3a). Table 2 is a list of the expressions

TABLE 2. EXPRESSIONS FOR $f(\frac{1}{2})$

2:(5.2) Sequence	$f(\frac{1}{2})$
1) N(n) $n = 2, 3, \ldots$	$f(\frac{1}{2}) = \frac{2^{n-1}}{D(2^{n+1} - 1)} [1 + n(2D - 1)]$
2) N(n)S $n = 2, 3, \ldots$	$f(\frac{1}{2}) = \frac{2^{n}}{D(2^{n+2} - 1)} [1 + (n - 1)(2D - 1) + \frac{D}{2^{n-1}}]$
3a) N(n)SN $n = 2, 3, \ldots$	$f(\frac{1}{2}) = \frac{2^{n+1}}{D(2^{n+3} - 1)} [1 + n(2D - 1) + \frac{D}{2^{n}}]$
3b) N(n)SS $n = 3, 4, \ldots$	$f(\frac{1}{2}) = \frac{2^{n+1}}{D(2^{n+3} - 1)} [1 + (n - 2)(2D - 1) + \frac{D}{2^{n-2}}]$
4a) N(n)SNN $n = 3, 4, \ldots$	$f(\frac{1}{2}) = \frac{2^{n+2}}{D(2^{n+4} - 1)} [1 + (n + 1)(2D - 1)]$
4b) N(n)SNS $n = 2, 3, \ldots$	$f(\frac{1}{2}) = \frac{2^{n+2}}{D(2^{n+4} - 1)} [1 + (n - 1)(2D - 1) + \frac{5}{2^{n+1}} \cdot D]$
4c) N(n)SSN $n = 3, 4, \ldots$	$f(\frac{1}{2}) = \frac{2^{n+2}}{D(2^{n+4} - 1)} [1 + (n - 1)(2D - 1) + \frac{5}{2^{n+1}} \cdot D]$
4d) N(n)SSS $n = 4, 5, \ldots$	$f(\frac{1}{2}) = \frac{2^{n+2}}{D(2^{n+4} - 1)} [1 + (n - 3)(2D - 1) + \frac{3}{2^{n-1}} \cdot D]$

for $f(\frac{1}{2})$ for all subintervals of I_n, n = 2, 3, ..., in which there are n + 4 discontinuity pairs or less; the interval of D to which each applies is given by Table 1.

In Table 3, expressions derived for Table 2 are used to evaluate <S(D)>, using equation (1.4), and with n = 2, 3, 4, 5. One interesting property of <S(D)> in this interval is the appearance of "plateaus" for certain intervals of D (e.g., N(2)S, N(3)SS, N(4)SSS, ...). The characteristic which identifies plateau regions is evident from examination of expressions for $f(\frac{1}{2})$; any sequence which involves a total of m_1 normalizations and m_2 subtractions yields an expression of the form:

$$f(\tfrac{1}{2}) = \frac{2^{T-2}}{D(2^T - 1)} [1 + (m_1 - m_2)(2D - 1) + \theta \cdot D] \qquad (3.4)$$

where θ is a constant. (Equation (3.1) can be used to confirm readily that (3.4) is true for the general case:

a) If the step required to generate x_{K+1}, given x_K, is a subtraction, then $x_K > 2D - \frac{1}{2}$ and the $a_K \cdot x_K^*$ product in (3.1) is of the form

$$\left[\frac{2D - 2^{K-1}(2D - 1) + \ldots}{2^{K-1}}\right] \cdot \frac{2^{T-2}}{D(2^T - 1)}$$

b) Analogously, if the kth step is a normalization, then x_K must be less than $\frac{1}{2}$, and the $a_K \cdot x_K^*$ product contributes a positive multiple of 2D - 1.)

Therefore, when m_1 exceeds m_2 by (exactly) one, (3.4) yields a constant. Conversely, $f(\frac{1}{2})$, and therefore <S(D)>, can be constant over an interval of D only when $m_1 - m_2 = 1$. Since nothing has been assumed regarding the value of

TABLE 3. EXAMPLE EXPRESSIONS FOR <S>, USING TABLE 2 AND EQUATION 3:(1.4)

Sequence in 2:(5.2)	<S>			
	n = 2	n = 3	n = 4	n = 5
1) N(n)	$\frac{7D}{2 - D}$	$\frac{15D}{8 - 9D}$	$\frac{31D}{24 - 33D}$	$\frac{63D}{64 - 97D}$
2) N(n)S	3	$\frac{31D}{8 - 3D}$	$\frac{63D}{32 - 35D}$	$\frac{127D}{96 - 131D}$
3a) N(n)SN	$\frac{31D}{8 - 3D}$	$\frac{63D}{32 - 35D}$	$\frac{127D}{96 - 131D}$	$\frac{255D}{256 - 387D}$
3b) N(n)SS	------	$\frac{63}{23}$	$\frac{127D}{32 - 9D}$	$\frac{255D}{128 - 137D}$
4a) N(n)SNN	------	$\frac{127D}{96 - 129D}$	$\frac{255D}{256 - 385D}$	$\frac{511D}{640 - 1025D}$
4b) N(n)SNS	3	$\frac{127D}{32 - 11D}$	$\frac{255D}{128 - 139D}$	$\frac{511D}{384 - 523D}$
4c) N(n)SSN	------	$\frac{127D}{32 - 11D}$	$\frac{255D}{128 - 139D}$	$\frac{511D}{384 - 523D}$
4d) N(n)SSS	------	---------	$\frac{255}{103}$	$\frac{511D}{128 - 25D}$

D, all solution subintervals in the $\frac{3}{5} < D < \frac{3}{4}$ region must correspond to sequences for which this difference is unity. A second consequence is that F(x) must have an even number of discontinuity pairs for intervals of D in which <S(D)> is constant, since the number of operations in 2:(5.2) is odd if $m_1 - m_2 = 1$.

In principle, Table 3 could be extended indefinitely to obtain algebraic expressions for <S(D)> for as great a fraction of the $.5 < D < .6$ interval as we desire, although the size of solution intervals decreases rapidly with larger numbers of discontinuities and are particularly minute in the lower portion of this interval of D. (See Table 1 for n = 3, 4, and 5.) Alternatively, by evaluating <S(D)> numerically, for a sufficiently fine grid of points in this interval, a graphical representation can be obtained to within any desired accuracy.

Figure 8 was obtained with the aid of ILLIAC II, using increments of .001 for $.5 < D < .6$. The discontinuity arguments, x^*_i, were determined for $i \leq 30$. By rewriting equation (3.1) as:

$$f(\tfrac{1}{2}) = \frac{2^{T-2}}{D(2^T - 1)} \left[1 + \sum_{i=0}^{T-2} \left\{ (\tfrac{1}{2})^i \cdot x^*_{i+1} \right\} \right] \qquad (3.5)$$

we observe that the maximum error in $f(\frac{1}{2})$ which can occur as a result of $T \leq 30$, where, in fact, T may be infinite, is of the order of 2^{-30}.

Using increments of 10^{-5} and the trapezoidal rule, a numerical approximation to the integral

$$\int_{.5}^{.6} <S(D)> dD$$

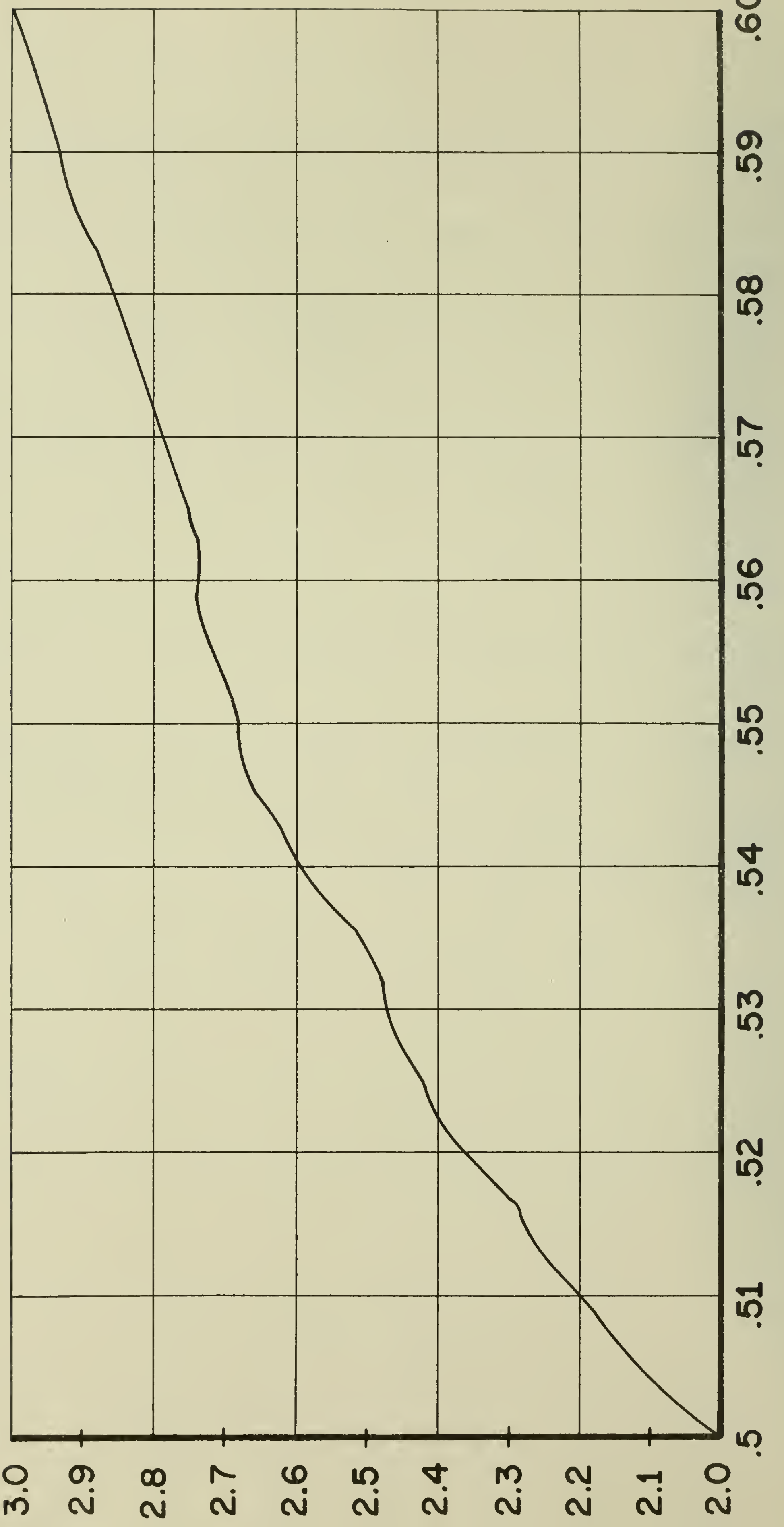


FIGURE 8: <S(D)> vs. D FOR .5 <D<.6

was found to be $2.61917 \pm 5 \times 10^{-6}$. This corresponds to the average value of $\langle S \rangle$ for D in this range, if all values of D are assumed equally likely. Using this result, together with 3:(1.4) and theorem 3.2b, the average $\langle S \rangle$ for the entire divisor range is 2.6170 for S-R-T division.

Because of the short computation time required per point (less than a millisecond), it was found practical to investigate in detail the behavior of $\langle S(D) \rangle$ for certain portions of the divisor range. The following are listed as empirical results:

a) The left limit of the plateau region in $\langle S \rangle$ which has $D = \frac{9}{16}$ as the right end point is $D = \frac{144}{257}$.

b) The above plateau is a local minimum in $\langle S \rangle$. Moving to the left of $D = \frac{144}{257}$, $\langle S \rangle$ has numerous local maxima, each of which is a plateau region. The limiting maxima is a point, viz., $D = \frac{24}{43}$, at which $\langle S \rangle = \frac{96}{35}$. While the difference of (approximately) 0.01 in the values of $\langle S \rangle$ at $D = \frac{24}{43}$ compared to the $\frac{144}{257} < D < \frac{9}{16}$ plateau may have little practical significance, this behavior is sufficiently contrary to intuition to warrant future investigation. Noteworthy is the much larger peak which Freiman's simulation of the so-called .75-1.0-1.5 method yields for D near $\frac{9}{16}$.

c) The nonmonotomic behavior of $\langle S \rangle$ described in (b) occurs again near $D = \frac{17}{32}$, and a plausible conjecture is that it will be found near all $D = \frac{1}{2} + (\frac{1}{2})^n$, $n = 4, 5, \ldots$, due to the similarities in these neighborhoods noted in Chapter 2.

4. SUMMARY AND CONCLUSIONS

4.1 Summary

Treating digital division as a Markov process in order to provide a basis for evaluating and comparing division algorithms was the contribution of Dr. C. V. Freiman. If the set of dividends encountered is assumed to have some degree of randomness, so that it may be approximated by a piecewise uniform probability distribution, then, as demonstrated in Chapter 1, the distribution function describing partial remainder magnitudes approaches a limit which is determined only by the divisor, D, and in particular, is independent of both the dividend distribution and the step numbers in the decision process. This stationary distribution is itself piecewise uniform, and the intervals of uniformity are the states of the Markov chain. However, a straightforward application of the theory of finite Markov chains in the case of S-R-T division was found impracticable, for the following reasons:

a) For S-R-T division, the number of discontinuities in the partial remainder stationary density function becomes indefinitely large in an infinite number of distinct neighborhoods in the divisor range, as demonstrated in Chapter 2.

b) When solving for the probabilities of the individual states, or intervals, the number of unknowns in the linear equation set must at least equal, and, in general, will exceed the number of discontinuities in the density function.

c) Solving one set of linear equations yields the partial remainder distribution for only one value of the divisor.

As an alternative to the finite Markov chain approach, a functional relation was derived which was shown to be a necessary and sufficient condition for the stationary density function (2:(1.5)). By considering the Fourier series representation of a periodic extension of the density function, and applying the functional relation just mentioned together with the knowledge that the total variation (per period) is bounded, it is proved that the density functions for all values of D must satisfy an important symmetry relation (Section 2.3). This symmetry is used to derive a number of other properties which lead to a straightforward algorithm (2:(5.2)) for determining all discontinuity arguments in the density function, and an equation for the size of the respective jumps (2:(5.6)). The limiting distributions being sought may be expressed in terms of these jumps. The solutions thereby obtained apply for intervals in the range of D, and by grouping these intervals on the basis of the sequence of steps used in 2:(5.2), infinite families of intervals are formed. Properties of the array of solution intervals are derived by considering, among other things, the relations which must exist between adjacent intervals.

While the partial remainder stationary density function is interesting in itself, a practical application lies in evaluating the shift average, $<S>$, which is the average number of quotient bits generated per iteration; this problem is considered in Chapter 3. Since the shift average is equal to the mean recurrence time of partial remainders which exceed 0.5 in magnitude, $<S>$ is found to be simply related to the partial remainder distribution function (3:(1.4)). Properties derived in Chapter 2 enable proving that $<S> = 3$ for the entire interval, $0.6 < D < 0.75$, a result which was previously not obtainable due to the complex behavior of the density function for D near 0.6. For $0.5 < D < 0.6$, the complicated behavior of the partial remainder density function is reflected in $<S>$ (Table 3), and relations derived in Chapter 3 are used to evaluate $<S>$ with aid of ILLIAC II, for D in that interval (see Figure 8).

These results, together with equation 3:(2.1), the expression for $<S>$ when $\frac{3}{4} < D < 1$, satisfy a goal defined at the outset: determination of the shift average for the entire range of divisors.

4.2 Conclusions and Areas for Future Investigation

In recent years there have been a number of efforts to better understand the digital division process when more than a minimum number of divisor multiples are available, and to apply this insight in formulating algorithms which reduce the number of nonzero quotient digits. Usually the avenue of approach is to consider bit patterns which can occur, then to select a set of rules which provides best performance in each example. This is often an iterative process.

Knowledge of stationary distributions of partial remainders for the various classes of division algorithms provides a valuable theoretical foundation in this search for optimum algorithms. Thus, for example, the shift average obtained for S-R-T division is applicable to an entire class of what might be called single-threshold division processes, which can be described as follows:

a) the divisor multiples available are: mD, 0, and -mD, where $m \geq 1$,

b) there is a positive number, K, such that $K \leq |mD| < 2K$ is true, and

c) K, which in the general case is a function of D, is used as the threshold level with which each R_i is compared to decide whether or not the divisor is to be used in forming R_{i+1}.

The abscissas in the <S(D)> versus D relation derived in Chapter 3 may be regarded as $\frac{m}{2K}$ D rather than D. Metze's modification of S-R-T division uses this change of scale effected by varying K to take advantage of the maximum which <S> was found to assume when $0.6 < D < 0.75$ is true.

An important question then is: can any similar generalizations be made to optimize multiple-threshold algorithm, perhaps using a property analogous to the relative independence of m and K in the single-threshold case? (Only the inequalities in (b) above must be satisfied.) In other words, when more than one (nonzero) multiple of the divisor magnitude is assumed to exist, to what extent can the solution for a specific algorithm be generalized in order to aid in formulating an optimum one?

Derivation of partial remainder distributions for any useful two-threshold algorithms (MacSorley's paper includes examples of such algorithms) is complicated by the fact that a number of ranges of the divisor must be analyzed individually, since the thresholds vary with D. However, the observations that a) the piecewise constant functions which describe the stationary density can be formed by locating discontinuity arguments as an ordered process, starting with discontinuities in the defining equation, and b) the size of jumps at successive arguments in this process form a geometric progression, should expedite these derivations.

In addition to considering binary division with further redundancy added, there is, of course, the problem of considering higher radices.

A second area where further work is needed is the modification of the <S(D)> versus D relation to account for the fact that, in parallel computers, shifting R_i an arbitrary number of places in one step is not possible. One approach to this problem would be that outlined by equations 3:(1.1), (1.2), and (1.3), with the index, m, in (1.1) confined to a finite range corresponding to the shift paths available.

Finally, a more thorough understanding of some peculiarities in the stationary distribution of R_i for S-R-T division would be interesting if not useful. A few of the questions which remain are:

a) Can the nonmonotonic behavior of $<S(D)>$ for $0.5 < D < 0.6$ be related to divisor bit configurations? The fact that "bad examples," i.e., cases for which S-R-T type division is clearly nonoptimal, often involve divisors of the form $\frac{1}{2} + (\frac{1}{2})^n$, and that the local minima in $<S(D)>$, which are plateau regions, include these points, indicates an intuitive explanation for these minima and what determined their end points may not be difficult.

b) What significance can be attached to the accumulation points of solution intervals, i.e., neighborhoods of D which lead to distributions of R_i with arbitrarily large numbers of discontinuities?

BIBLIOGRAPHY

1. A. T. Bharuca-Reid, Elements of the Theory of Markov Processes and Their Applications, Chapt. 1, McGraw-Hill; 1960

2. J. Cocke and D. W. Sweeney, "High Speed Arithmetic in a Parallel Device," IBM internal report, February 1957

3. W. Feller, An Introduction to Probability Theory and Its Applications, second edition, Chapt. 14, 15, John Wiley and Sons; 1960

4. C. V. Freiman, "Statistical Analysis of Certain Binary Division Algorithms," Proceedings of the IRE, vol. 49, No. 1, pp. 91-103; January 1961

5. J. G. Kemeny, H. Mirkhil, J. L. Snell, G. L. Thompson, Finite Mathematical Structures, Chapt. 6, Prentice-Hall, Inc.; 1959

6. O. L. MacSorley, "High Speed Arithmetic in Binary Computers," Proceedings of the IRE, vol. 49, No. 1, pp. 67-91; January 1961

7. G. A. Metze, "A Class of Binary Divisions Yielding Minimally Represented Quotients," IRE Transactions on Electronic Computers, vol. EC-11, pp. 761-4; December 1962.

8. Penhollow, J. O., "A Study of Arithmetic Recoding with Applications in Multiplication and Division," University of Illinois Digital Computer Laboratory. (Doctoral Dissertation); 1962

9. R. K. Richards, Arithmetic Operations in Digital Computers, D. Van Nostrand Co.; 1955

10. J. E. Robertson, "A New Class of Digital Division Methods," IRE Transactions on Electronic Computers, vol. EC-7, pp. 218-222; September 1958

11. T. D. Tocher, "Techniques of Multiplication and Division for Automatic Binary Computers," Quart. J. Mech. Applied Math., vol. XI, pt. 3, pp. 364-384; 1958

12. J. B. Wilson and R. S. Ledley, "An Algorithm for Rapid Binary Division," IRE Transactions on Electronic Computers, vol. EC-10, pp. 662-670; December 1961

13. A Zygmund, Trigonometric Series, vol. I, Chapts. 1, 2, Cambridge; 1959